

博士論文

機械学習を利用した顕著性物体検出と
その重要度推定に関する研究

提出日 令和3年1月29日
所属 長岡技術科学大学大学院
工学研究科 博士後期課程
情報・制御工学 専攻
学籍番号 15501484

梅木 陽

目次

第 1 章	まえがき	1
1.1	研究背景及び目的	1
第 2 章	基礎理論	11
2.1	概要	11
2.2	顕著性検出	12
2.3	顕著性物体検出	12
2.4	重要度付き顕著性物体検出	13
2.5	セマンティックセグメンテーション	14
2.6	インスタンスセグメンテーション	14
2.7	画像リターゲティング	15
2.8	第 2 章のまとめ	16
第 3 章	顕著性物体の高精度化手法	18
3.1	関連研究:顕著性物体の高精度化手法	18
3.2	従来法:顕著性物体の高精度化手法	24
3.3	提案法:顕著性物体の高精度化手法	30
3.4	シミュレーション:顕著性物体の高精度化手法	37
3.5	第 3 章のまとめ	40
第 4 章	顕著性物体の重要度推定	47
4.1	関連手法:顕著性物体の重要度推定	47
4.2	重要度について:顕著性物体の重要度推定	52
4.3	提案法:顕著性物体の重要度推定	60
4.4	シミュレーション:顕著性物体の重要度推定	64
第 5 章	結論	79

謝辭	82
參考文獻	83
研究業績一覽	88

目次

1	従来法による顕著性物体マップ [1]	5
2	顕著性物体検出と重要度付き顕著性物体検出の結果を用いた画像リターゲティング	8
3	入力画像に対する各手法の出力結果	14
4	セマンティックセグメンテーションとインスタンスセグメンテーションの差異	15
5	画像リターゲティングの結果	17
6	スーパーピクセル分割による結果	19
7	CIELab 色空間	23
8	従来法の概略図	25
9	提案法概略図	32
10	従来法と提案法の初期ラベル比較	33
11	初期顕著性増加の様子	37
12	実験 1 の MAE	43
13	実験 1 の F-measure	44
14	実験 2 の MAE	44
15	実験 2 の F-measure	45
16	実験 1 の結果画像	45
17	実験 2 の結果画像	46
18	(a) 入力画像と (b) 重要度付き顕著性物体検出の真値	49
19	複数物体画像における顕著性物体検出と重要度付き顕著性物体検出を用いた画像リターゲティングの結果	55
20	重要度付き顕著性物体検出の提案手法概略図	56
21	提案 CNN のアーキテクチャ	57

22	α に対する提案評価指標の変化：横軸は α 縦軸は評価値を表す	73
23	σ に対する提案評価指標の変化：横軸は σ 縦軸は評価値を表す	74
24	PASCAL-S データセットにおける結果画像 [2]	75
25	SALICON ベースのデータセットにおける結果画像 [3,4]	76
26	重要度推定比較:スピアマン相関係数	77
27	重要度推定比較:提案評価指標	78

表目次

1	各章間の関連	10
2	実験 1 の MAE	42
3	実験 1 の F-measure	42
4	実験 2 の MAE	42
5	実験 2 の F-measure	43
6	PASCAL-S データセット内の画像が含む物体数とその割合 [2]	49
7	PASCAL-S データセットにおける重要度推定法別の評価値 [2]	57
8	提案 CNN アーキテクチャの構造	70
9	任意のベクトルに対する順位相関と提案評価指標のスコア	71
10	DUT データセットにおける F-measure [5]	71
11	PASCAL-S データセットにおける F-measure [2]	71
12	SALICON ベースのデータセットにおける F-measure [3,4].	72
13	PASCAL-S データセットにおける重要度推定のスコア [2]	72
14	SALICON ベースのデータセットにおける重要度推定のスコア [3,4]	72

第 1 章

まえがき

1.1 研究背景及び目的

本論文では，機械学習を用いた顕著性物体検出手法とそれら顕著性物体の重要度推定手法を提案する．近年，SNS などの普及により大量の画像を扱う機会が増え，画像処理の自動化が望まれている．中でも，画像リターゲットと呼ばれる画像処理技術が盛んに研究されておりインターネットブラウザ，SNS，テレビ中継などに利用されている．画像リターゲットは，画像内に与えられた重要な物体を残したまま違和感なく画像を変形する技術であり主に重要な物体領域ではないピクセルを様々な方法で縮小，拡大することでこれを実現している [6–9]．そのため，高精度なリターゲットのためには重要物体領域の正確な検出は必要不可欠である．

画像リターゲティングのために重要な物体を検出する手法の一つに、顕著性物体検出がある [2, 10–29]. 顕著性物体検出は、画像から人間が興味を持つ物体として、人間の視点が集まる物体だけを検出する技術である。画像リターゲティングにおいて顕著性物体検出を利用しない場合、一般物体検出などにより物体領域を重要な領域として検出する方法が考えられる。しかし、その場合は主要物体以外の物体も検出された結果、全ての物体を縮小後の画像に保持する必要があるため主要物体の形状が大きく変化しリターゲティングの精度が劣化してしまう。そのため、視点の集まる物体のみを検出する顕著性物体検出が必要とされる。

より自由度の高い画像リターゲティングを行うために、顕著性物体の重要度推定が必要となる。本研究では、より視点の集まる物体がより重要であるとして、重要度を定義した。これは、人間の視点が集まる度合いに応じた離散的な値である。顕著性物体に重要度が設定されていることで、最も重要な物体だけを残したりターゲティングや、二番目までを残したりターゲティングなどユーザーの要求に応じた自由度の高いリターゲティングが可能となる。そのため、より高度な画像リターゲティングのためには、重要度付きの顕著性物体検出が必要となる。

そこで、本論文では顕著性物体検出の高精度化手法と、顕著性物体の重要度推定手法の二つについて提案を行う。顕著性物体検出の高精度化手法では、従来法で複数物体が写った画像において誤検出が起きる問題を解決した。これは、初期顕著性の検出の提案と競合

型伝播の提案によって実現される。次に、データセットや妥当な評価指標が少なくほとんど研究の行われてこなかった重要度推定について本研究で実現した。これは、重要度の詳細な定義づけ、妥当な評価指標の検討、そして重要度推定の提案手法で構成されている。

はじめに、顕著性物体検出の高精度化手法について説明する。近年、様々な顕著性物体検出手法が提案されており最近では明らかに顕著な局所領域を検出しそれを基に物体を検出する二段階の手法が提案されている [1, 12, 16]。特に、明らかに顕著な領域を検出しそれらとの類似度を基に画像全体の顕著性を推定する手法や、一般的に画像端は非顕著な領域と判断し色情報によるそれらとの類似度と距離を用いたマルコフ確率場による推定方法などが良好な結果を示している。 [1, 12]

それら二段階的な手法の中で、伝播を用いた手法が特に良好な成果を挙げている。 [1] この手法では、まず、局所領域の色および構造情報を基に物体である可能性を算出する **Objectness** を用いて明らかに顕著な領域を検出し、またマルコフ確率場による手法と同様に画像端領域を非顕著領域と設定する。次に、それら初期の顕著性情報を画像全体に反映させるため、各局所領域の色情報から計算された隣接領域間の類似度を用いて伝播処理を行う。それにより、画像全体で物体及び背景ごとに分けられた顕著性物体マップが生成される。特に、この伝播方法は初期顕著性を逐次的に伝播するため、緩やかに変化する領域でも初期顕著性が伝播しやすく顕著性情報が距離によって減衰することなく物体全体に与えられ、他手法と比較して検出物体の一部が非顕著となる問題が低減されている。

しかし、この手法は初期顕著性の決定が粗野であるために図 1 (c) 上段のような不正確な検出結果を示す場合がある。従来の初期顕著性検出法では、前述したように画面端が非顕著な初期顕著性として設定されていた。これは画像端が背景である可能性が高いことに基づいているため有効な手法ではあるが、物体が多数ある複雑な画像の場合には画像端以外にも背景領域が存在することが多くそれらを取りこぼしてしまう問題がある。伝播を用いた手法では、高コントラスト領域や顕著性物体が隔離された領域へは伝播が及ばないため、複雑な画像を検出対象とした場合に顕著性の誤検出が起こる。それを解決するためには、画像全体に初期顕著性を分布させなければならない。

また、伝播手法は初期顕著性の波及効果が高いため顕著性情報を物体全体に与えられるが、一方で類似度がやや高い領域間では図 1 (c) のように誤検出が起きる。伝播手法は、明らかに顕著および非顕著であるという初期顕著性をそれぞれ別々に伝播させ二つの顕著性物体マップを作成したのちに、それらを加重結合によって平均化して最終的な顕著性物体マップを生成している。しかし、加重結合では複数の顕著性物体マップの内どれか一つにでも誤検出が起きている場合にはその誤検出を抑制する手段がない。そのため、良好な初期顕著性が設定された場合でもどちらか片方でも誤検出した場合にはその誤検出が最終結果に残留し、精度が低下してしまう問題がある。

本研究では、その問題を解決するために SIFT (Scale-invariant Feature Transform) 特徴量によって検出された特徴点の密度と機械学習を利用した離散的な初期顕著性と競合型伝



図1 従来法による顕著性物体マップ [1]

播を用いた高精度な顕著性物体検出手法を提案する。提案法は、誤った初期顕著性による誤検出を低減するために、SIFT 特徴点密度を用いて画像全体に初期顕著性を分布させる。背景領域は、一般的に輝度変化や色変化が少なくなだらかな領域である。SIFT 特徴量は、輝度や色の変化が大きい領域に多数現れるため SIFT 特徴点の密度によって背景領域を検出できる。これにより、初期顕著性を画像全体に分布させ孤立領域の発生を防ぐ。また、提案法は誤検出が残留することによる精度の低下を防ぐために、顕著および非顕著初期顕著性の伝播を一つのマップ上で同時に行う。一枚マップ上での同時伝播により、片方ずつ伝播していた場合に誤検出が残留する問題が消え精度の低下を抑制できる。また、初期顕著性が画像内に偏っている場合も誤検出要因となる。そこで、初期顕著性を画像全体にまばらに配置するため初期顕著性の逐次的増加を行う。これは、確実に顕著、非顕著な小領

域を伝播により付加された顕著性値と初期顕著性領域との類似度を基に探索することで行う。これらにより、提案法は顕著性の誤検出を抑え物体全体が顕著であると検出される高精度な顕著性物体検出を実現する。

次に、顕著性物体の重要度推定手法について説明する。顕著性物体検出に関する研究は数多くあり、複雑なデータセットを用いた実験などで高い検出精度を示す等優良な成果を挙げている。しかし、これらの手法は顕著性物体を検出するのみにとどまっており、それら顕著性物体の重要度を推定するには至っていない。これは、重要度に関する詳細な定義が行われていないことや妥当な評価指標が存在しないこと、データセットが少ないことが大きく影響している。しかし、重要度推定は、図2に示すように画像リターゲティングにとって重要である。ここで、図2の(a)は入力画像、(b)と(c)は顕著性物体検出と順位付き顕著性物体検出での理想的な顕著性物体マップ、(d)と(e)は、それぞれ[8]手法を使って(b)と(c)の結果から(a)をリターゲティングしたものである。(b)では、白と黒の領域はそれぞれ顕著な領域と非顕著な領域を表し、(c)では、白、灰色、および黒の領域はそれぞれ第1の顕著な領域、第2の顕著な領域、および非顕著な領域を表す。(d)では、(b)に示す顕著性物体検出によって椅子と犬に同じ重要度が与えられているため、(a)で最も重要なオブジェクトである犬の一部が切り取られてしまっている。対照的に、(c)は椅子と犬に別々の重要度が付加されており犬の方が重要であることがわかっているため、(e)の画像では椅子よりもより重要な犬を優先して残した画像リターゲティン

グが実現されている。図 2 は拡張の利点のひとつを示しており、画像リターゲティングだけでなく、コンテンツに応じた画像コーディングや画像表現などにも利用できる可能性がある。

そのため、本研究では重要度付きの顕著性物体検出を提案する。そして、その定義、利用先、および評価について議論することにより、顕著性物体検出を改良した新しいテーマとして重要度付き顕著性物体検出を提案する。まず、いくつかの例を用いて重要度付き顕著性物体検出の定義とその利用先について説明および定義を行う。本論文では、先行研究の分析から重要度の定義を議論し、重要度は人間の視点が集まるほどに高くなり N 段階で表されるものであるとして定義する。また、顕著性物体検出の利用先として挙げられる画像リターゲティングにおいて重要度が果たす役割について議論し、重要度付与の妥当性を議論する。この議論に基づいて、本論文では重要度付き顕著性物体検出の評価手法とデータセットの作成法を提示する。評価手法については、二乗誤差とケンドールの順位相関係数に基づく評価指標を提案し、データセットについては顕著性検出とインスタンスセグメンテーションの検出結果に基づいて提案する。最後に、機械学習とインスタンスセグメンテーションに基づく重要度付き顕著性物体検出を提案する。

実験では、顕著性物体の推定精度と重要度の推定精度について最新の従来法と比較を行っている。顕著性物体の推定精度に関する実験では、提案法は従来法に比べ初期顕著性の誤りと加重結合による誤検出を低減し、より高精度な顕著性物体検出を実現している。



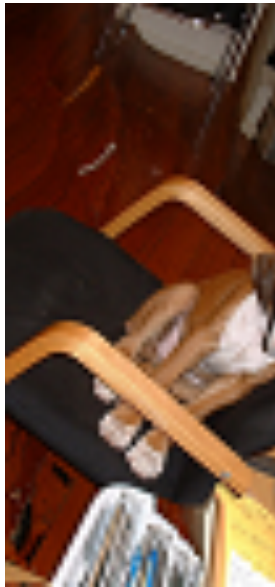
(a) 入力画像



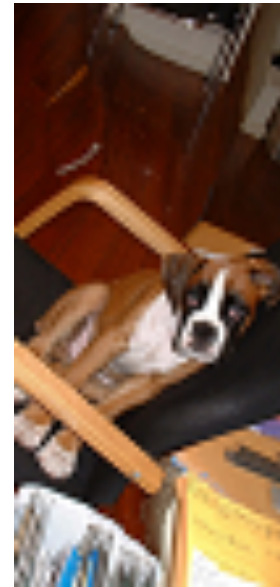
(b) 顕著性物体検出の検出結果



(c) 重要度付き顕著性物体検出の検出結果



(d) (b) を基にしたリターゲティング結果



(e) (c) を基にしたリターゲティング結果

図2 顕著性物体検出と重要度付き顕著性物体検出の結果を用いた画像リターゲティング

また、顕著性物体の重要度推定に関する実験では順位だけを推定していた従来の手法に比べて提案法は順位の評価と実値での評価で共に優良な成果を占めており、定量的かつ知覚

的に優れていることを示した。

本論文は，以下のように 5 章構成になっており，各章は表 1 のように関連している．本論文では，顕著性物体検出の高精度化を目的として画像全体にまばらに配置された初期顕著性を用いて競合型伝播を行うことで顕著性物体の誤検出を低減する手法を提案している．また，顕著性物体の重要度推定を目的として従来考慮されていなかった重要度について定義と評価指標を議論し，視点データと物体領域を同時に学習した機械学習モデルによる重要度推定手法を提案している．第 1 章では，顕著性物体検出を目的とする従来研究を概観し物体検出との関連をまとめ，本研究の目的と範囲，各章間の関係を説明している．第 2 章では，第 3 章以降に記述される共通事項として，顕著性検出，顕著性物体検出，セマンティックセグメンテーション，およびインスタンスセグメンテーション，それら手法の主要な利用先として画像リターゲティングについて概要を簡潔にまとめている．第 3 章では，離散的初期顕著性を用いた競合型伝播による顕著性物体検出の高精度化手法について述べる．初期顕著性をそれぞれ HDCT 特徴量と SIFT 特徴点を用いて推定することで正確な初期顕著性の推定を提案している．次に，競合型ラベル伝播を用いて検出した二つの初期顕著性を同時に画像全体へ伝播させていき競合させる．結果，加重結合で起きていた誤検出の残留が低減され顕著性物体の検出精度が改善されることを実験的に確認している．第 4 章では，重要度付き顕著性物体検出にがついて述べる．第 4 章では，現状実現していなかった重要度付き顕著性物体検出について詳細な議論を行ったことと，議論に基づ

表1 各章間の関連

1章-各章	共通部 差異	各章の概説を述べる 詳細は各章に記載
2章-3章	共通部 差異	顕著性物体検出の従来法を概説 提案手法との具体的な差を述べる
2章-4章	共通部 差異	物体領域抽出や顕著性検出の従来手法を概説 提案手法との差を述べる，提案法の一部として利用
3章-4章	共通部 差異	顕著性物体検出を行う 3章：高精度化が目的，4章：重要度推定が目的
5章-各章	共通部 差異	各章の結論を述べる 詳細は各章に記載

いて重要度付き顕著性物体検出の提案手法を示したことの二つの貢献が述べられている。

はじめに，顕著性物体検出の発展として，物体ごとに多値の重要度を与える重要度付き顕著性物体検出を定義する。また，これを利用したアプリケーションの拡張性や評価指標についても調査，提案を行い，本提案が画像処理アプリケーションに与える貢献を明確にする。この議論に基づき，視点データと物体領域を同時に学習した機械学習モデルによる重要度推定手法を提案する。結果，従来考慮されていなかった重要度について推定出来ることを実験的に確認している。第5章では，本論文の結論並びに今後の課題について述べている。本論文により得られた成果は，画像リターゲティングにおける主要物体の変形を低減するのに利用できるほか，ユーザーごとの要求に合わせた画像リターゲティングを実現できる。よって本論文は工学上および工業上貢献するところが大きい。

第 2 章

基礎理論

2.1 概要

本節では、本研究に関連する手法として顕著性検出，顕著性物体検出，セマンティックセグメンテーション，およびインスタンスセグメンテーションの従来法について説明し，それら手法の主要な利用先として画像リターゲットングについて説明する．はじめに，顕著性検出と視線推定についておおまかに説明しその目的について説明する．次に，顕著性物体検出について説明しそれらの主要な手法について説明する．また，順位付き顕著性物体検について説明し，顕著性物体検出との違いについて述べる．さらに，セマンティックセグメンテーションとインスタンスセグメンテーションの違いについて説明し，近年の手法について説明を行う．最後に，画像リターゲットングについて説明し画像リターゲット

ングに前述した方法を適用した場合についての差異を述べる。

2.2 顕著性検出

顕著性検出は人間の興味領域，つまり視線を推定する手法である [30–38]。従来の顕著性検出手法では，コントラストの高さ，特徴的な色，独特な方向などの特徴を利用する [30]。Harel らは，グラフベースのアルゴリズムを利用していくつかの特徴に基づいて活性化マップを計算し，それらを組み合わせて一つの顕著性マップを生成する方法を提案した [31]。近年では，畳み込みニューラルネットワーク（CNN）を利用した手法が数多く提案されており，多数の画像とそれに対応する視点情報を学習データとしてトレーニングし大域的かつ複雑な特徴を効果的に抽出する [35, 36]。これら顕著性検出手法は，図 3 (b) のように人間の視点を正確に推定することを目的としており，物体領域を輪郭に沿って検出することはできない。

2.3 顕著性物体検出

顕著性物体検出は，図 3 (c) のように物体領域とそれらが顕著性を持つかどうかを同時に推定する技術である [2, 10–26, 28, 29]。従来の顕著性物体検出手法は，伝播アルゴリズムを使用する [1, 12, 23]。それらは，隣接するピクセル間の色の類似性とマルコフ吸収

確率に基づいて、顕著性情報と背景情報を繰り返し処理によって伝播する。ただし、物体境界に沿って不正確な結果が生成される場合がある。近年では、完全畳み込みネットワーク（FCN）アーキテクチャに基づく手法が提案され、良好な成果を挙げている。Liuらは顕著性物体を推定するために粗い情報と詳細な情報を同時に推定する Deep hierarchical saliency ネットワークを提案した [22]。また、再帰的ネットワークを採用して顕著性を持つピクセルがどの程度接続しているかを推定する手法が提案されている [17]。

2.4 重要度付き顕著性物体検出

主要なデータセットの一つには物体の重要度を含むものが存在するが、従来の手法のほとんどが顕著性を持つか持たないかだけの二値の推定結果を出力する。PASCAL-S データセットは、図 3 (d) 物体領域の輪郭と $[0, 255]$ の整数の顕著性値を持っている [2]。しかし、従来の顕著性物体検出手法はこの重要度を無視し、顕著性物体を物体領域に沿って正確に検出することを目標としている。これは、顕著性物体を物体領域通りに検出することは難しい課題であるため、ほとんどの顕著性物体検出の研究において重要度に関して将来的な作業としているからである。

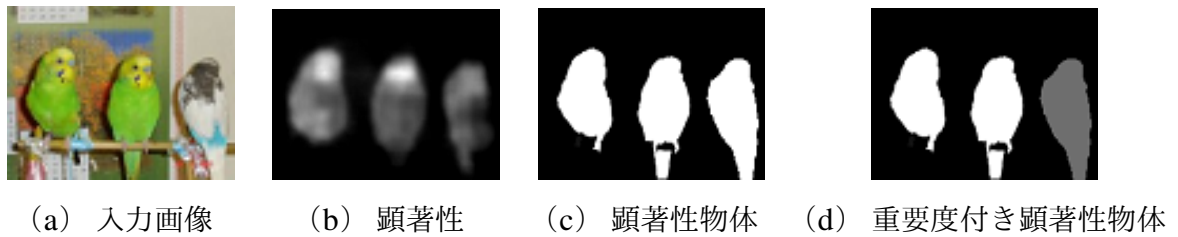


図3 入力画像に対する各手法の出力結果

2.5 セマンティックセグメンテーション

セマンティックセグメンテーションは、図4 (b) のように人間、木、車など、ピクセルが所属するカテゴリを識別する手法である [39–41]. 従来のセマンティックセグメンテーションでは、輪郭検出と勾配のヒストグラムによって識別を行っていた [39]. 近年、画期的なアプローチである FCN が画像領域の正確に検出する方法として注目を集め、様々な手法で利用されている [40]. ただし、セマンティックセグメンテーションでは、同一カテゴリの物体は一つとして認識されるため、同じカテゴリの物体を別々に識別することはできない.

2.6 インスタンスセグメンテーション

インスタンスセグメンテーションはセマンティックセグメンテーションから派生したものである [42,43]. これは、図4 (c) のように物体のクラスだけではなく、それぞれの物体領域も識別できる. 基本的なインスタンスセグメンテーション手法は、FCN を使用し



図4 セマンティックセグメンテーションとインスタンスセグメンテーションの差異

て、それぞれが一つの物体を含む小さな窓を検出する [43]. 別の方法では、リカレントネットワークのアーキテクチャを利用して、以前の検出結果に基づいて物体領域を繰り返し検出する [42]. インスタンスセグメンテーションと顕著性物体検出は同様に物体領域を推定する手法であるが、インスタンスセグメンテーションは物体の重要度を考慮しないため、それぞれの目的は異なっている.

2.7 画像リターゲティング

画像リターゲティングは、図5に示すように画像の重要な領域を保存したまま違和感なく画像を変形する技術である [6–9]. 画像の重要な領域は主に物体領域であるため、前述したセマンティックセグメンテーション、インスタンスセグメンテーション、顕著性物体検出などが重要領域抽出に利用される. [8]に代表されるように、画像リターゲティングは破棄してもいいピクセル群を重要領域を基に探索する. この時、セマンティックセグメンテーションやインスタンスセグメンテーションのように物体領域をすべて抽出してし

まうとより重要な物体が欠損し不要な物体が保存されてしまう場合がある。一方で顕著性物体検出をこの重要領域に利用した場合，顕著性物体検出は人間の視点が集まる重要な物体のみを検出し不要な物体を検出しないため重要な物体のみを保存したりターゲティングが可能となる。そのため，画像リターゲティングには顕著性物体検出がより適しているといえる。また，この顕著性物体それぞれに多値の重要度が付いている場合，画像リターゲティングはユーザーの要求に合わせて一番重要な物体のみを保存したりターゲティングや二番目までの物体を保存したりターゲティングなど，重度の高いリターゲティングが可能になる。

2.8 第 2 章のまとめ

第 2 章では，本研究に関連する手法として顕著性検出，顕著性物体検出，セマンティックセグメンテーション，およびインスタンスセグメンテーションの従来法について述べた。また，上記 4 つの差異について小領域を検出するものと物体を輪郭通りに検出するもの，さらにすべての物体を検出するかどうかの観点から述べた。さらに，これら 4 つの手法を用いてリターゲティングを行った場合にどのような差が出るかを述べた。

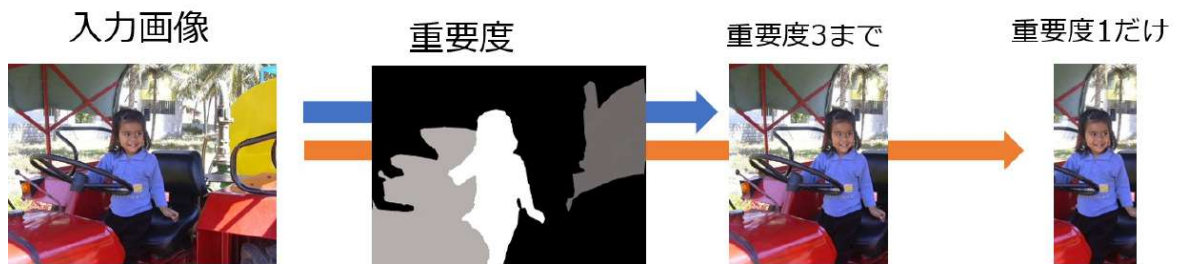


図5 画像リターゲティングの結果

第 3 章

顕著性物体の高精度化手法

3.1 関連研究:顕著性物体の高精度化手法

3.1.1 スーパーピクセル

スーパーピクセル分割は、図 6 に示したような隣接の類似画素をひとまとめにしたスーパーピクセルと呼ばれる重複のない小領域に分割する技術であり、画像処理の分野において広く用いられている [44]. スーパーピクセル分割は類似した画素をひとまとめにし、その領域に対して一度に処理を行うことを可能にするため高速な処理を実現することができる。また、大まかに物体領域を分割することが出来るため物体検出への利用や、小領域ごとの特徴ベクトル抽出にも利用される。

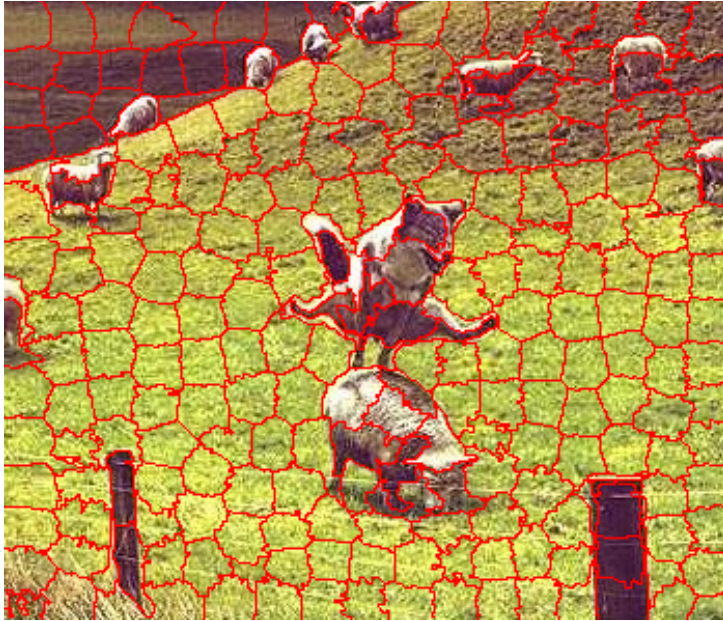


図6 スーパーピクセル分割による結果

3.1.2 特徴抽出

HOG 特徴量

HOG (Histogram of Gradients) 特徴量は、局所領域における輝度の勾配方向をヒストグラム化した照明変化に不変な特徴量である [44,45]。HOG 特徴量は、画像を $n \times n$ 画素の格子状に区切ったセルごとにヒストグラムを求め、さらに $m \times m$ セルをブロックと呼ばれるひとまとめにして正規化しブロックごとに特徴量を抽出する。特徴量の次元 dim は以下の式で求められる。

$$dim = \left(\frac{W}{n^2} - m^2 + 1\right) \times \left(\frac{H}{n^2} - m^2 + 1\right) \times m^4 \times O \quad (1)$$

ここで、 W と H はそれぞれ画像の横幅と縦幅を表しており、 n^2 、 m^2 はそれぞれセルサイズとブロックサイズを表している。また、 O はヒストグラムに使用する勾配方向数である。また、座標 (x, y) の勾配強度 $magnitude(x, y)$ と勾配方向 $gradient(x, y)$ はそれぞれ次式で求められる。

$$\begin{aligned} magnitude(x, y) &= \sqrt{grad_{hol}^2 + grad_{ver}^2} \\ gradient(x, y) &= \tan^{-1}\left(\frac{grad_{ver}}{grad_{hol}}\right) \end{aligned} \quad (2)$$

ここで、 $grad_{hol}$ と $grad_{ver}$ はそれぞれ水平方向と垂直方向の勾配強度を表し、 $grad_{hol}(x, y) = I(x + 1, y) - I(x - 1, y)$ 、 $grad_{ver}(x, y) = I(x, y + 1) - I(x, y - 1)$ 、 $I(x, y)$ は画像 I の座標 (x, y) におけるグレースケールでの画素値を表している。

SIFT 特徴量

SIFT (Scale-Invariant Feature Transform) 特徴量は、照明変化や回転、拡大縮小に頑強な特徴量であり画像マッチングや物体追跡、パノラマ写真合成などに利用されている [46, 47]。その手法は、まず画像上の特徴的な場所を表す特徴点を検出し特徴点からの輝度勾配を基に特徴ベクトルを記述する。はじめに、特徴点は DoG (Difference of Gaussian) 処理を用いて極値探索をすることで検出する。DoG とは入力画像をガウシアンフィルタによって平滑化しその差分を取ることを指し、SIFT 特徴量ではその差分画像において極値をとる点を特徴点と定義している。次に、検出した特徴点の周囲で輝度勾配

のヒストグラムを計算し各特徴点から特徴ベクトルを生成する。特徴点の周囲の輝度勾配は前述した HOG 特徴量とほぼ同様の方法で計算され 36 方向に勾配が求められる。この勾配の強度は特徴点のガウシアンフィルタで重み付けされるため、拡大縮小に不変な特徴量となっている。また、この特徴点の勾配を基準の方向とすることで回転に対しても強い特徴量となっている。

3.1.3 機械学習 (ランダムフォレスト)

多くの手法が、顕著性物体検出のために機械学習を利用している。本研究では、その中のうちの一つある Kim を利用する。この手法は、自然画像とその顕著性物体の真値から前述した HDCT (High-dimensional Color Transform) 特徴量と呼ばれる特徴量をスーパーピクセルごとに抽出し、学習を行う。そして、学習したモデルによって入力画像よりスーパーピクセルごとに領域が顕著性物体かそうでないかを検出する。学習にはランダムフォレストを利用し、特徴ベクトルを学習している。

3.1.4 CIELaB 色空間

本研究で利用している CIELab 色空間は、一般的に用いられる RGB や CMYK とは異なり人間の視覚に近似するように設計されている色空間である [48]。図 7 に示すように CIELab は明度を意味する次元 L と補色次元の a および b によって構成されており、 a は

赤と緑方向の色の強さを表し b は黄と青方向の強さを表す次元である。人間の目の感度はすべての色について同じではなく、RGB の数値上で差があっても目で判断できない領域が存在するため、RGB 空間では色差の大きな色とされる場合でも人間の目から見た場合には類似した色と判断する場合がある。しかし CIELab 色空間は RGB 上では色差の大きな色同士でも人の視覚と同様に似た色であると判定することが出来る。顕著性領域は、人間の目に興味深く感じる領域であるため人間の目と差異のある RGB より CIELab 色空間により精度の高い検出ができる。CIELab は、RGB から直接変換することはできず XYZ を経由して変換される。RGB から XYZ への変換は変換行列 \mathbf{M} を用いて以下の式で行う。

$$\begin{pmatrix} X \\ Y \\ Z \end{pmatrix} = \mathbf{M}^{-1} \begin{pmatrix} R \\ G \\ B \end{pmatrix} \quad (3)$$

ここで、

$$\mathbf{M} = \begin{bmatrix} 0.489 & 0.176 & 0 \\ 0.311 & 0.183 & 0.010 \\ 0.201 & 0.011 & 0.990 \end{bmatrix}. \quad (4)$$

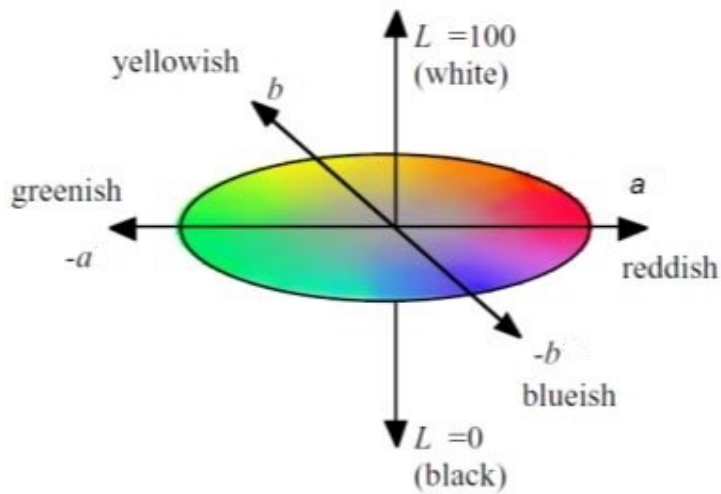


図7 CIE Lab 色空間

求められた $(X, Y, Z) \in ([0, 1], [0, 1], [0, 1])$ から CIE Lab への変換は以下の式で行う.

$$\begin{aligned}
 L &= 116 \times f(Y/Y_n) - 16 \\
 a &= 500 \times (f(X/X_n) - f(Y/Y_n)) \\
 b &= 200 \times (f(Y/Y_n) - f(Z/Z_n))
 \end{aligned}
 \tag{5}$$

ここで, X_n, Y_n, Z_n は白色点の値であり, $(X, Y, Z) = (1, 1, 1)$ である. また, f は以下

となる.

$$f(t) = \begin{cases} t^{\frac{1}{3}} & \text{if } t > (\frac{6}{29})^3 \\ \frac{(\frac{29}{3})^3 \times t + 16}{116} & \text{otherwise} \end{cases}
 \tag{6}$$

3.2 従来法：顕著性物体の高精度化手法

3.2.1 従来法概要

この節では、二段階的手法を用いた従来法の中でも特に優れた結果を有する伝播を基にした顕著性検出手法について述べる [1]. Li らが提案した伝播を基にした手法は、近年主流になっている二段階的な顕著性検出手法の中でも特に良好な検出結果を示す手法である。図 8 に従来法の概要図を示す。各画像は左から入力画像、初期顕著性、検出された顕著性物体、最終結果の順である。また、矢印は処理の流れを示している。その手法は、まず確実に非顕著な領域として画像端を背景初期顕著性に設定し、同時に確実に顕著な領域として Objectness によって物体と検出された領域を前景初期顕著性に設定する [49]。画像端を非顕著な領域として設定するのは、人間が撮影した自然画像の場合は撮影者が顕著な領域を画像端に配置することが非常に稀だからである。また、Objectness は高い物体認識精度を有するので多くの画像において物体の特徴領域を検出でき、従来法では前景初期顕著性検出に用いられている。次に、設定されたそれぞれの初期顕著性を伝播処理によって画像全体に広げ二つの顕著性物体検出結果を生成する。伝播は繰り返し処理によって行われ、領域の類似度に従って隣接した領域へ次々と初期顕著性が伝播される。最後に、伝播処理によってそれぞれの初期顕著性から推定された二つの顕著性物体検出結果を加重結

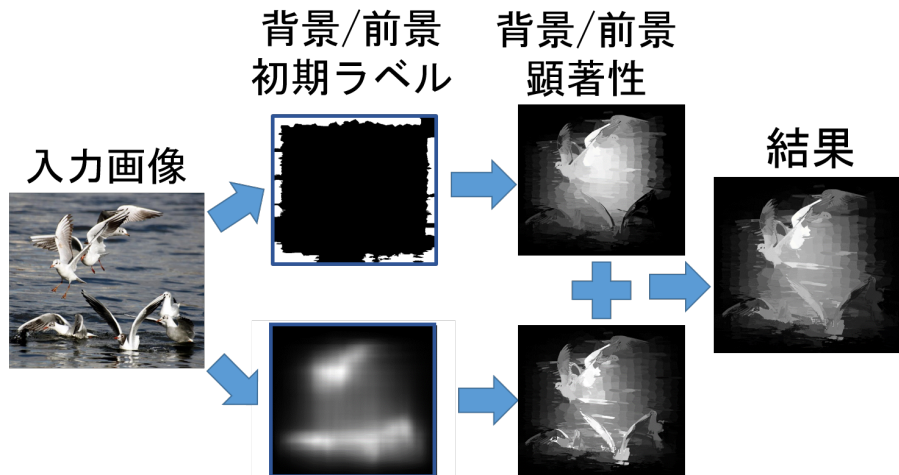


図8 従来法の概略図

合により足し合わせて最終的な顕著性物体検出結果を得る。

3.2.2 初期顕著性

初期顕著性を取得する手法は様々な形で提案されている。従来の顕著性検出手法を用いて推定された顕著性マップを基に、非常に高いもしくは低い顕著性を有する領域をそれぞれ顕著もしくは非顕著な局所領域と定義する方法がある [30,31,50]。ただし、その方法では手法が特徴的な物体を顕著な領域と定義していないため、色変化の少ない物体を顕著領域としない場合や背景にあるテクスチャ領域を顕著領域と誤検出することが多く発生するため有効な手法ではない。Sun らは、人が撮影した一般的な画像の場合に重要な物体は画像の中央に配置されるとし、画像端付近の領域は非顕著な領域であると定義している [12]。また、Li らは、同様の理由から画像端付近を非顕著領域と定義するとともに、局所領域の色および構造情報を基に物体である可能性の度合いを測る *Objectness* を用いて

その値が非常に高い局所領域を顕著領域と定義している [1]. さらに Kim らは, 三つの色空間における値とヒストグラム値, 近傍との輝度値差, HOG 特徴量などから構成される 71 次元の特徴量ベクトルを定義しそれを基にランダムフォレストを用いて初期顕著性情報の推定を実現している [51]. それらの手法は, 近年の顕著性検出の目的に沿って定義されており判定の条件が厳しく設定されているため, 顕著もしくは非顕著であると推定される領域数は少ないものの正確な検出を達成できている.

3.2.3 初期顕著性の伝播

本節では, 従来法において提案された伝播手法を示す. その手法は, 画素単位ではなく局所領域単位で定義されておりスーパーピクセルに適用される. いくつかのスーパーピクセルに与えられた初期顕著性は, 隣接スーパーピクセル間の類似度を基に繰り返し処理によって伝播され, 一つ前の結果と比較して変化がなくなったら繰り返し処理を終了する. 最終的な顕著性物体検出結果は $[0, 1]$ 内の実数値で構成され必要に応じて閾値処理により二値化される.

$\mathbf{v}_t \in \mathbb{R}^N$ を繰り返し処理が t 回目の検出結果とすると以下で算出される [48].

$$\mathbf{v}_t = F_{\mathbf{v}_0}(\mathbf{A}\mathbf{v}_{t-1}) \quad (7)$$

ここで, \mathbf{A} は親和行列であり $\mathbf{A} = \mathbf{D}^{-1}\mathbf{W}$, $\mathbf{D} = \text{diag}(q_1, q_2, \dots, q_N)$, $q_i = \sum_{j=1}^N s_{ij}$,

diag は対角行列， \mathbf{W} は (i, j) 要素が s_{ij} である対称行列をそれぞれ表し， N はスーパーピクセルの総数を表している．初期顕著性ベクトル \mathbf{v}_0 は，初期顕著性が与えられたスーパーピクセルの値を 1 にしてそれ以外を 0 に設定されたものである． $F_{\mathbf{v}_0}$ は，初期のスーパーピクセルのインデックスをリセットする関数であり以下となる．

$$[F_{\mathbf{v}_0}(\mathbf{x})]_i = \begin{cases} 1 & \text{if } [\mathbf{v}_0]_i = 1 \\ [\mathbf{x}]_i & \text{otherwise} \end{cases} \quad (8)$$

ここで， $[\mathbf{x}]_i$ は \mathbf{x} の i 番目要素を表す．また， s_{ij} は i 番目と j 番目のスーパーピクセルの類似度を示しており以下のように定義される．

$$s_{ij} = \begin{cases} \exp\left(-\frac{d(\mathbf{f}_i, \mathbf{f}_j)}{\sigma^2}\right) & \text{if } j \in \mathcal{N}(i) \text{ or } i, j \in \mathcal{B} \\ 0 & \text{if } i = j \text{ or otherwise} \end{cases} \quad (9)$$

ここで， d はユークリッド距離を， \mathbf{f}_i は i 番目スーパーピクセルの CIELab 色空間上の色差値のみの特徴ベクトル， σ は類似度を制御するパラメータ， $\mathcal{N}(i)$ は i 番目スーパーピクセルの近傍にあるスーパーピクセルのインデックス集合， \mathcal{B} は画像端にあるスーパーピクセルのインデックス集合をそれぞれ表している．

3.2.4 HDCT 特徴量

いくつかの顕著性検出手法は機械学習を用いて検出を行っており、本論文でも Kim らが提案した機械学習手法を利用する [14,16,52]. その手法は、大量の自然画像とその視点データより機械学習アルゴリズムを用いて事前学習モデルを生成し顕著性物体を検出する。本論文で用いる手法は、それぞれのスーパーピクセルから HDCT 特徴量と呼ばれる特徴量を算出し機械学習アルゴリズムに入力する。その特徴量は、それぞれのスーパーピクセル内の HDCT 特徴量, HOG 特徴量, 特異値より構成される [16,44,45,53]. 機械学習アルゴリズムにはランダムフォレストを用いている。顕著性検出を行う際は、特徴ベクトルを入力画像のスーパーピクセルごとに計算し事前学習モデルによって分類する。HDCT 特徴量は、画素の座標と勾配, 色, 色ヒストグラム, 色コントラストの特徴を多数の色空間上で組み合わせたものである。この特徴量は、いくつかの特徴から定義されそれぞれ位置特徴, 色特徴, テクスチャ特徴, 色ヒストグラム特徴, 色コントラスト特徴と呼ばれている。位置特徴は画像上での各スーパーピクセルの中心座標である。色特徴は、RGB と CIELab, HSV 色空間上での画素値を平均したものである。テクスチャ特徴は各スーパーピクセルの画素数と HOG 特徴量, 特異値から構成されている [53]. ヒストグラム特徴は、あるスーパーピクセルとその周囲のスーパーピクセルとの RGB と CIELab, ヒ

ストグラム値を基にカイ二乗距離によって以下のように計算される。

$$H_i = \sum_{j=1}^N \sum_{k=1}^B \frac{(h_{ik} - h_{jk})^2}{(h_{ik} + h_{jk})} \quad (10)$$

ここで H_i は i 番目のスーパーピクセルにおけるヒストグラム特徴であり、 N と B はそれぞれスーパーピクセルの総数とヒストグラムのビン数を表している。さらに、 h_{ik} は i 番目のスーパーピクセル中の k 番目のビンにおける RGB ヒストグラム値である。同様にして HSV と CIELab 空間上でのヒストグラム特徴も計算することが出来る。コントラスト特徴は、グローバルコントラストとローカルコントラスト、それぞれの色特徴の分布によって構成されている。グローバルコントラストとローカルコントラストはあるスーパーピクセルとそれ以外のスーパーピクセル間の距離により算出される。 G_i と L_i をそれぞれ i 番目のスーパーピクセルとのグローバルコントラストとローカルコントラストとすると、

$$\begin{aligned} G_i &= \sum_{j=1}^N (c_i - c_j)^2, \\ L_i &= \sum_{j=1}^N \omega_{ij} (c_i - c_j)^2, \\ \omega_{ij} &= \frac{1}{Z_i} \exp\left(-\frac{1}{2\sigma_p^2} \|\mathbf{p}_i - \mathbf{p}_j\|_2^2\right), \end{aligned} \quad (11)$$

となる。ここで、 c_i は i 番目のスーパーピクセルにおける色特徴の一要素を表している。

Z_i は正規化係数を表していて、 $\mathbf{p}_i \in [0, 1] \times [0, 1]$ で定義されるのは i 番目のスーパーピクセルの位置特徴である。 σ_p はフリーパラメータで、距離に対する重み付けを行って

いる。

3.3 提案法：顕著性物体の高精度化手法

3.3.1 概要

本論文では、従来法での一段階目である初期顕著性検出と二段階目である伝播においてそれぞれ発生する初期顕著性の誤検出と加重結合によって誤検出が残留する問題を解決するために SIFT 特徴点密度を用いた背景初期顕著性検出と競合型伝播を提案する。提案法の背景初期顕著性検出は、画面端を背景初期顕著性とすることに起因する誤検出を低減するために、SIFT 特徴点密度を利用し画像全体に背景初期顕著性を分布させる。従来法は、自然画像の場合に顕著な物体は画面端に配置されることが少ないので背景初期顕著性を画面端に設定していた。そのため、画像中央部に大きな背景領域が存在する場合は画像端の初期顕著性が届かず不明な領域として顕著性が誤検出される事がしばしば発生する。提案法は、伝播が及ばないせいで起きる誤検出を防ぐことができる。

提案法の競合型伝播は、従来法で問題であった加重結合によって誤検出が残存する問題に対して、相反するラベルを同時に伝播させることで誤検出を低減する。従来法では、多数の物体が存在する複雑な画像に対して初期顕著性を片方ずつ伝播し、加重結合を行うことで加重結合前に起きていた誤検出が最終結果まで残存してしまう問題があった。それ

は、初期顕著性が伝播される際に物体境界外へ顕著性情報が流出するために発生すると考えられ、類似度を基に伝播処理を行う手法では起こり得る問題である。しかし、従来法の加重結合では発生した誤検出を低減する方法がないため、片方にでも誤検出が発生した場合に最終結果にその誤検出が残留してしまう問題を抱えていた。そこで提案法では、相反するラベルを同時に伝播することで誤検出の発生を低減し加重結合で誤検出が残存する問題を解決する。さらに、繰り返し伝播の過程で前景背景ともに信用できる領域を探索し初期顕著性として順次追加して誤検出の発生を抑制する。図 9 に提案法の概略図を示す。ここで、 mod は余剰関数であり t は繰り返し回数を表している。また、 n は初期顕著性の追加回数をコントロールするパラメータであり繰り返し回数 t が n の倍数に達するたびに初期顕著性を追加する。はじめに、入力画像をスーパーピクセルに分割し 3.3.2 節で示す初期顕著性検出によって設定した初期顕著性情報のうち顕著な領域は 1 と非顕著な領域は -1 、それ以外の領域は 0 として従来法と同様にベクトル化する。それに対して 3.3.3 節で示した伝播を適用し終了判定を行う。本手法では、繰り返し n 回毎に 3.3.3 節で示す初期顕著性の追加処理を行う。

3.3.2 初期顕著性検出

提案手法では、SIFT 特徴点密度による背景初期ラベル検出と HDCT と機械学習を用いた前景初期顕著性検出により、画像中央部に顕著性が誤検出される問題を解決する。前述

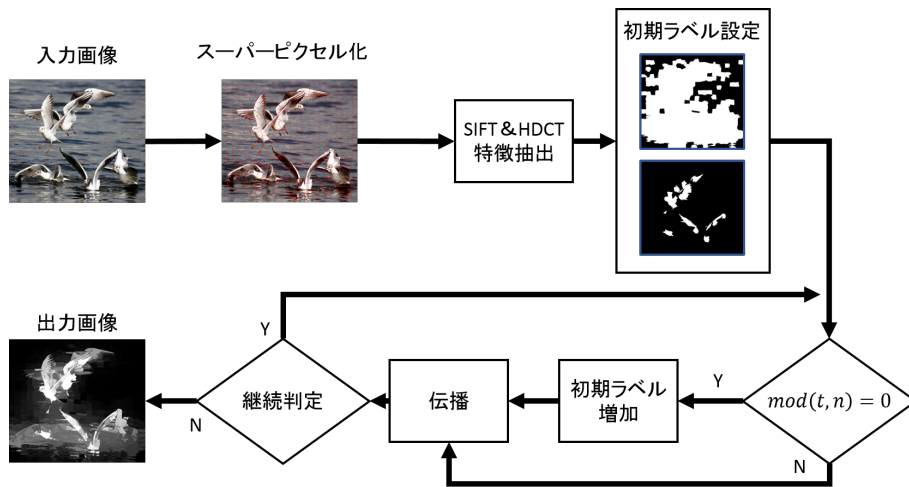
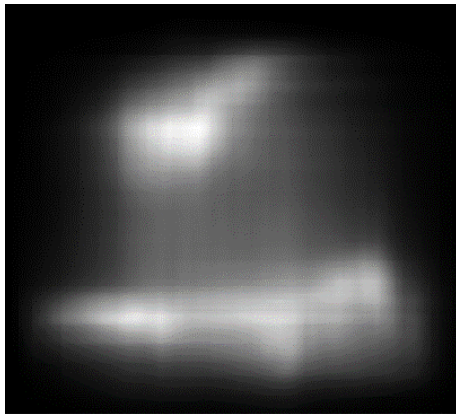


図9 提案法概略図

したとおり，伝播を基にした手法において初期顕著性は不正確な顕著性を避けるように画像全体に設定されなければならない．前景初期顕著性は，HDCT を基にランダムフォレスト (RF) 手法によって設定される [51]．その手法は，Objectness を利用した従来法と比較して信頼できる前景エリアをより正確に検出でき，さらに Objectness では検出できないような小さな物体や画像端付近にある物体に対しても検出精度が高い．次に，背景初期顕著性を SIFT 特徴点密度が低い場所に設定する．3.1.2 節で示したとおり，SIFT は DoG を基に特徴点を検出するため物体境界や輝度変化の大きな場所に現れる．よって，背景領域は輝度変化の少ない領域やぼやけた領域であると定義されるため，SIFT 特徴点の少ない領域は背景と設定できる．これら二つの初期顕著性検出手法により，前景と背景の初期顕著性は隔離された領域にも設定され伝播が及ばないことによる顕著性の誤検出を回避できる．



(a) 入力画像



(b) 従来法前景初期ラベル



(c) 提案法前景初期ラベル



(d) 従来法背景初期ラベル



(e) 提案法背景初期ラベル

図 10 従来法と提案法の初期ラベル比較

前景初期顕著性

前景初期顕著性を事前学習モデルによって設定する。そのモデルは、多くの自然画像とその顕著性マップを用いて HDCT 特徴量を基にランダムフォレストによって学習される。

入力画像の各スーパーピクセルに対して前景らしさを $[0, 1]$ で数値化し, そのスコアがパラメータ θ_f を越えた領域は前景初期顕著性として設定される. θ_f はフリーパラメータであり本手法では 0.75 に設定した. 図 10 (b) に従来法によって設定された前景初期顕著性を, (c) に提案法によって設定された前景初期顕著性を示す. 提案法では, 従来法と比較して背景領域への前景初期顕著性の付加を低減することができているのがわかる.

背景初期顕著性

背景初期顕著性を SIFT 特徴点密度を基に設定する. はじめに, 入力画像より SIFT 特徴量を用いて特徴点を検出する. g_i を i 番目のスーパーピクセルの特徴点密度とし, \bar{g} を全スーパーピクセルにおける特徴点密度の平均値とする. そして, $g_i \leq \bar{g} \times \theta_b$ であるときに i 番目のスーパーピクセルを背景初期顕著性として設定する. ここで, θ_b は背景初期顕著性を設定する特徴点密度を決めるフリーパラメータであり, 本手法では 0.2 に設定する. 図 10 (d) に従来法によって設定された背景初期顕著性を, (e) に提案法によって設定された背景初期顕著性を示す. 提案法では, 従来の背景初期顕著性と比較して画像全体に背景初期顕著性が分布し孤立した領域を低減できている.

3.3.3 提案伝播法

概要

提案する競合型伝播は，物体境界を越えての伝播を低減するために相反する顕著性情報を同時に伝播させる．前述したとおり，従来法では片方の初期顕著性情報のみを伝播させるせいで，誤検出が発生しやすくまた加重結合によってその誤検出が解消されず残存していた．それに対して，相反する顕著性情報を同時に伝播することで誤検出とその残存を抑えられる．さらに，繰り返し伝播の過程で顕著性マップと近傍の初期顕著性を基に信用できる領域を探索し初期顕著性への追加を行う．これにより，初期顕著性から離れた領域への伝播が活発になるとともに，画像全体にまばらに初期顕著性を拡散できる．相反する顕著性の同時伝播において誤検出が発生するのは，誤った顕著性情報が伝播されているからであり，初期顕著性の追加により該当領域に正しい初期顕著性が増加すれば誤検出が低減され，顕著性物体の検出精度が向上する．

競合型伝播

競合型伝播は、従来法では前景と背景で別々に行われていた伝播を同一マップ上で行うことで実現する。そのために、式 (7) で示した $F_{\mathbf{v}_0}$ を以下のように定義する。

$$[F_{\mathbf{v}_0}(\mathbf{x})]_i = \begin{cases} 1 & \text{if } [\mathbf{v}_0]_i = 1 \\ -1 & \text{if } [\mathbf{v}_0]_i = -1 \\ [\mathbf{x}]_i & \text{otherwise} \end{cases} \quad (12)$$

ここで、初期顕著性ベクトル \mathbf{v}_0 は背景初期顕著性が与えられたスーパーピクセルのインデックスを -1 、前景初期顕著性が与えられたスーパーピクセルのインデックスを 1 、それ以外を 0 に設定されたものとする。この初期顕著性ベクトルに式 (7) の処理を繰り返し適用する。最後に、生成された顕著性物体検出結果を $[0, 1]$ に正規化し出力する。

初期顕著性の追加

提案法では、繰り返し処理の過程で n 回毎に初期顕著性領域を追加する。式 (9) で示した類似度 s を用いて、 $\text{mod}(t, n) = 0$ での顕著性マップから初期顕著性が以下のように追加される。

$$[\mathbf{v}_0]_i = \begin{cases} 1 & \text{if } \tau_1 < [\mathbf{x}]_i \text{ and } \tau < s_{ij} \\ -1 & \text{if } \tau_2 > [\mathbf{x}]_i \text{ and } \tau < s_{ij} \end{cases} \quad j \in \mathbb{L} \quad (13)$$



図 11 初期顕著性増加の様子

ここで、 \mathcal{L} は i 番目スーパーピクセルの近傍に存在する初期顕著性のインデックスを表す。また、 τ , τ_1 , τ_2 は初期顕著性を追加するための閾値である。図 11 に初期顕著性追加の様子を示す。図中の (b) は提案法によって $t = 0$ 時に設定された背景初期顕著性であり、(c) は $t = 100$ 時の増加された初期顕著性を含む背景初期顕著性である。 $t = 0$ 時と比べて初期顕著性の増加が確認できる。ただし、初期顕著性を多く追加すると異なる顕著性情報境界で不均衡となり誤検出を引き起こす可能性があるため、この追加は頻繁には行わない。

3.4 シミュレーション：顕著性物体の高精度化手法

3.4.1 実験概要

本章では、提案法と従来法の顕著性検出を比較して実験を行う。MIT-300 データセットを用いて正しい顕著性物体マップの生成は 30 人の被験者による主観評価実験によっ

て行った [54]. また, 比較対象として Li らと Sun ら, Kim らの手法を用い, それぞれ LPS, MAP, HDCT と表記する [1, 12, 16]. また, 提案法の τ , τ_1 , τ_2 はそれぞれ $\tau = \tau_1 = -\tau_2 = 0.7$ であり, スーパーピクセルの分割数は全ての手法で $N = 500$ とする.

初期顕著性に関する結果と伝播に関する結果をそれぞれ得るために二つの条件で行う.

一つ目の実験は従来法と提案法の初期顕著性の検出精度を比較するため伝播方法は LPS で提案された方法を共通で用いている. 二つ目の実験は伝播性能を確認するために理想的な初期顕著性を用いて行う.

定量評価は, Mean Absolute Error (MAE) と F-measure の二つの指標で行った. MAE は精度予測評価指標の一つで, 予測値が正解から平均的にどの程度乖離しているかを示す指標で精度の悪さを示すため 0 に近い値が良い結果となる. MAE は以下の式で算出される.

$$\text{MAE} = \frac{1}{N} \sum_{i=1}^N |y_i - \hat{y}_i| \quad (14)$$

ここで y_i と \hat{y}_i はそれぞれ観測値と真値を表し, N はスーパーピクセルの総数を表している. F-measure は正確性と網羅性の総合的な評価に利用される指標であり, 適合率と再現率と呼ばれる評価指標から以下の式で算出される.

$$\text{F-measure} = \frac{2 \times \text{適合率} \times \text{再現率}}{\text{適合率} + \text{再現率}} \quad (15)$$

ここで、適合率は結果と真値がどれだけ合致しているかを示すものであり、

$$\text{適合率} = \frac{tp}{tp + fp} \quad (16)$$

によって求められる。また、再現率は結果として出てくるべきものがどの程度網羅されているかを示す指標であり、

$$\text{再現率} = \frac{tp}{tp + fn} \quad (17)$$

によって求められる。ここで、式中の tp , fp , fn はそれぞれ truepositive, false positive, false negative を表している。

3.4.2 実験結果

表 3.5 と 3.5 に MAE の数値結果を、表 3.5 と 3.5 に F-measure による数値結果をそれぞれ示す。表内の ‘Prop.’ は提案手法を ‘Average’ はそれぞれの手法を 20 枚の画像に適用した数値結果の平均値を示している。

図 12 は実験 1 における MAE の数値結果を示しており、提案手法の平均値が 0.053 と最も低い値であり従来法 LPS から改善されている。また、図 13 は実験 1 における F-measure の数値結果を示しており、提案法の平均値が 0.737 と最も高い数値であり従来法 LPS から改善されている。同様に図 14 と図 15 は実験 2 における MAE の数値結果と

F-measure の数値結果をそれぞれ示しており，MAE の結果が従来法 LPS が 0.209 に対して提案法 0.093 と改善し F-measure の結果も従来法 LPS が 0.578 に対して提案法が 0.650 と改善されている。

次に，視覚的評価を図 16 と図 17 に示す。図 16 では初期顕著性の改善によって不正確な顕著性が低減されていることがわかる。また，図 17 では加重結合によって残留していた誤検出が競合型伝播によって低減し顕著性物体の検出精度が向上しているのがわかる。

3.5 第3章のまとめ

第3章では，顕著性物体について概説し顕著性物体検出が画像処理の分野において有用な技術であることを述べた。3.1 節では，顕著性物体検出で主に用いられる CIELab 色空間，スーパーピクセル手法について解説した。さらに，従来法で用いられる HOG 特徴量と提案法で用いた SIFT 特徴量についても解説した。3.2 節では，従来の顕著性検出の概要と具体的な手法について述べた。また，従来法による顕著性検出画像を示し粗野な初期ラベル推定によって起きる問題と伝播後の加重結合によって不正確な顕著性が検出されることを指摘した。そして，従来法で利用されている HDCT 特徴による機械学習手法について述べた。3.3 節では，SIFT 特徴点密度を用いた背景初期ラベル検出及び HDCT による機械学習を用いた前景初期ラベル検出と競合型ラベル伝播について述べた。また，繰り

返し処理中の初期ラベルの増加について述べた。3.4節では、提案法によって生成された顕著性検出画像と入力画像の理想的な顕著性を MAE と F-measure, 視覚的評価によって比較した。定量的な評価では、従来法 LPS に対して提案手法は高い数値結果を示し、正確に顕著性が検出されていることを確認した。また、視覚的な評価においても従来法に比べて正確な初期顕著性設定と競合型伝播によって誤検出が低減され、提案法の有効性を認めることができた。

表2 実験1のMAE

	Prop.	MAP [12]	LPS [1]	HDCT [16]
小舟	0.022	0.058	0.037	0.028
ピエロ	0.035	0.125	0.126	0.077
灯台	0.031	0.080	0.110	0.070
商店	0.076	0.303	0.086	0.088
猿	0.122	0.265	0.188	0.123
Average	0.053	0.137	0.120	0.067

表3 実験1のF-measure

	Prop.	MAP [12]	LPS [1]	HDCT [16]
小舟	0.799	0.072	0.667	0.768
ピエロ	0.821	0	0.068	0.577
灯台	0.740	0.580	0.036	0.338
商店	0.659	0.345	0.644	0.499
猿	0.593	0.412	0.141	0.356
Average	0.737	0.449	0.302	0.592

表4 実験2のMAE

	Prop.	LPS [1]	MAP [12]	HDCT [16]
小舟	0.056	0.225	0.037	0.327
牛	0.068	0.216	0.126	0.206
音楽家	0.074	0.171	0.110	0.212
ピエロ	0.082	0.205	0.086	0.332
女性	0.183	0.227	0.086	0.293
Average	0.153	0.186	0.120	0.226

表5 実験2のF-measure

	Prop.	LPS [1]	MAP [12]	HDCT [16]
小舟	0.648	0.305	0.037	0.190
牛	0.899	0.622	0.126	0.655
音楽家	0.805	0.635	0.110	0.541
ピエロ	0.709	0.446	0.086	0.379
女性	0.729	0.658	0.086	0.504
Average	0.650	0.578	0.120	0.510

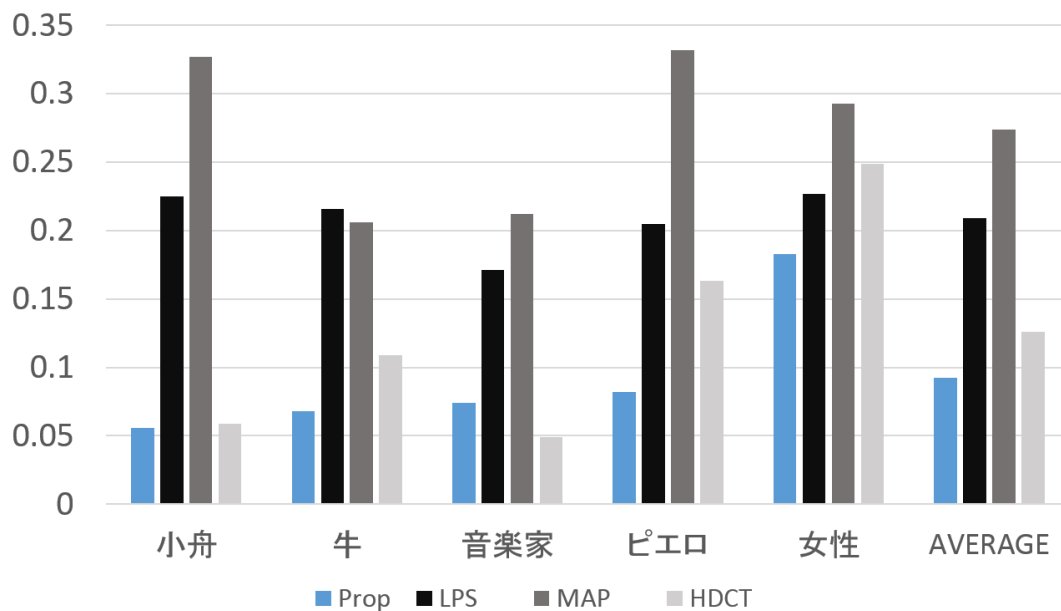


図12 実験1のMAE

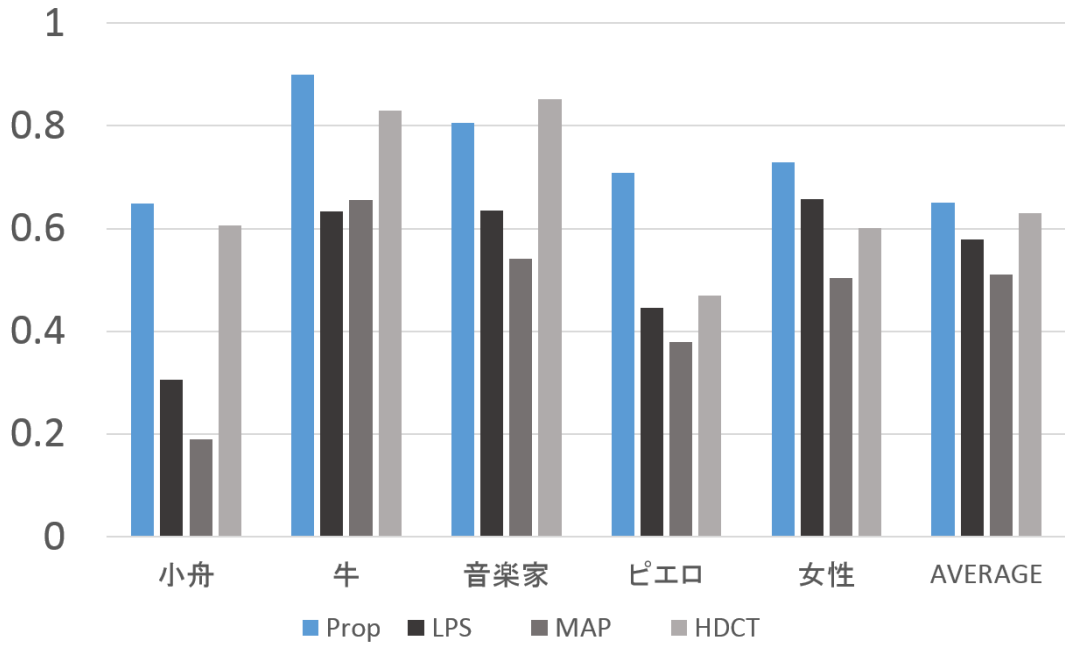


図13 実験1のF-measure

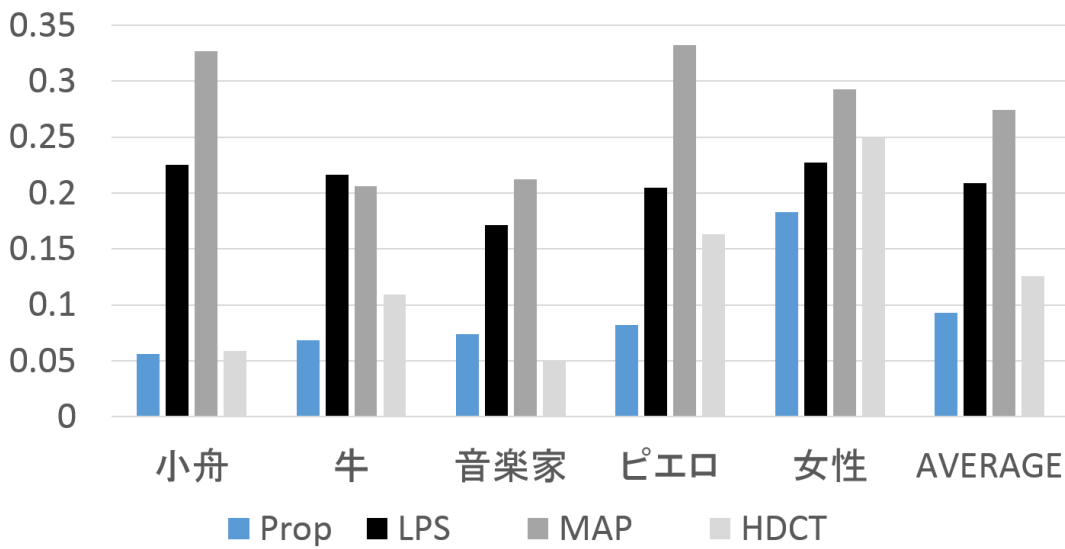


図14 実験2のMAE

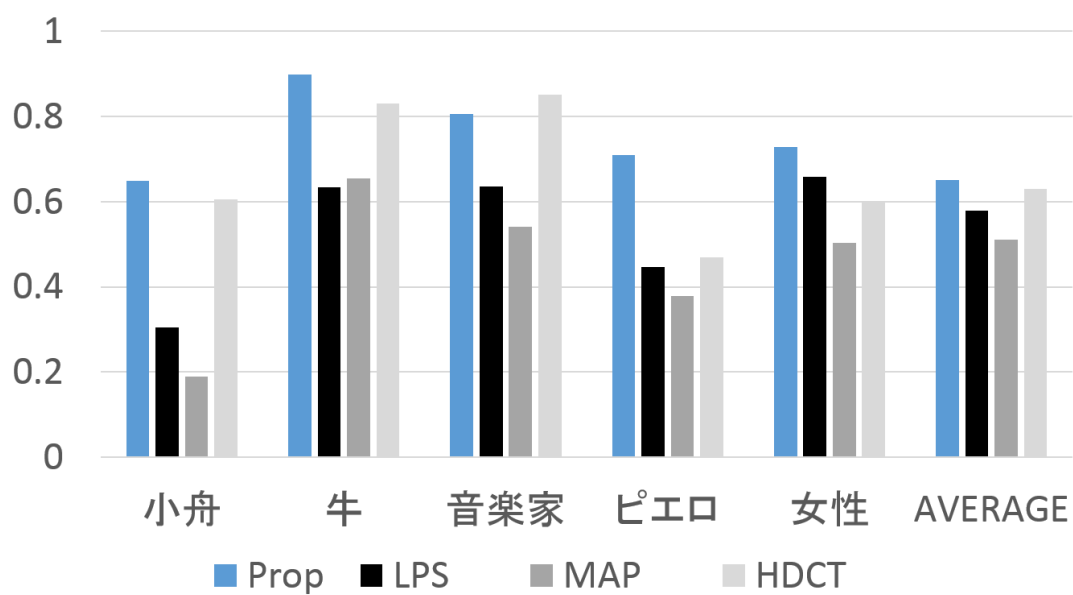


図 15 実験 2 の F-measure

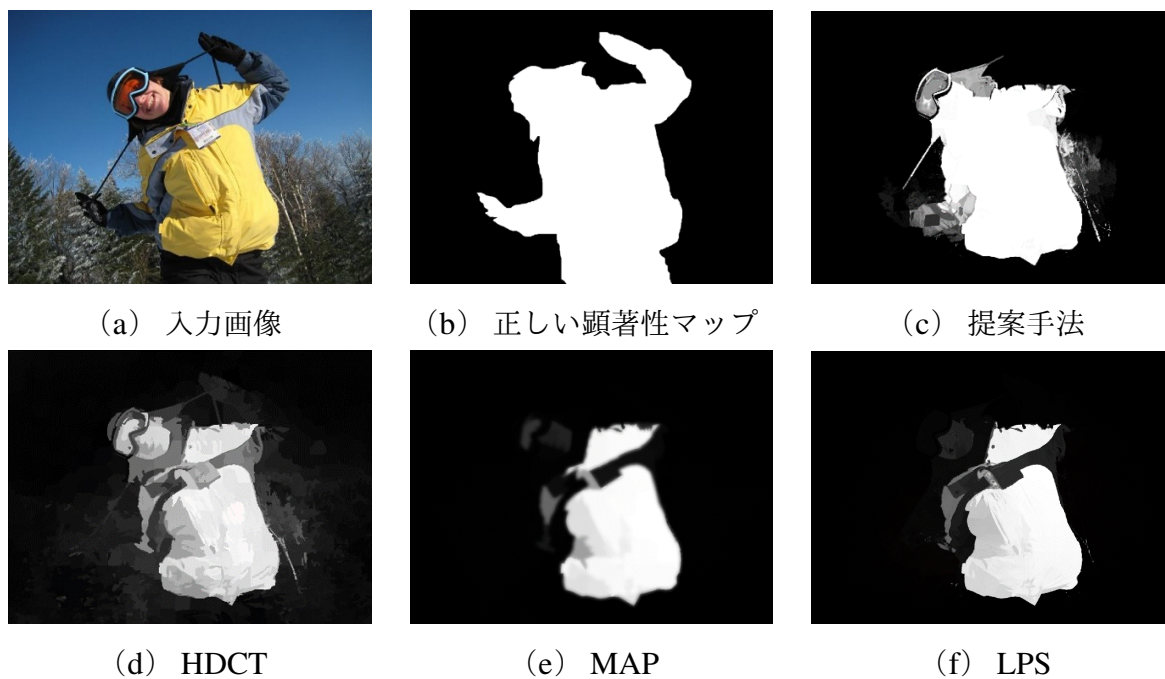


図 16 実験 1 の結果画像

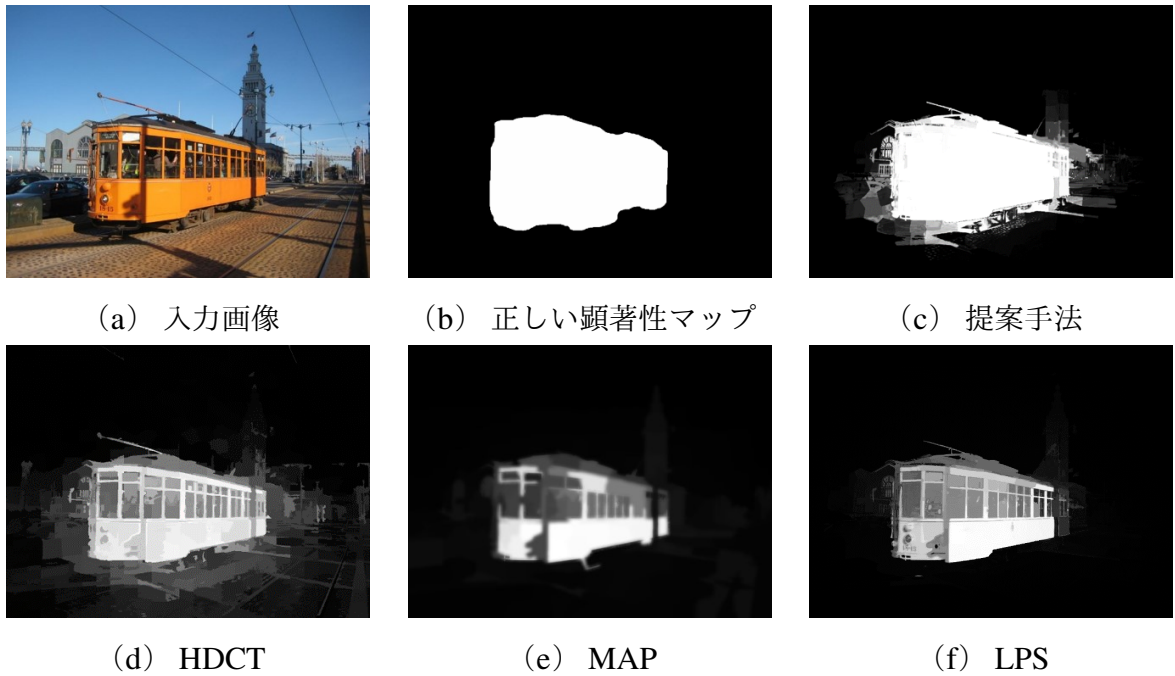


図17 実験2の結果画像

第 4 章

顕著性物体の重要度推定

4.1 関連手法：顕著性物体の重要度推定

4.1.1 PASCAL-S データセットの分析

PASCAL-S データセットは自然画像と視点データ，そしてそれらに基づいて段階的な値を与えられた顕著性物体の真値画像が含まれている [2]。PASCAL-S データセットの真値画像は段階的な値を持っているが，通常二値化されて顕著性物体の真値として利用される。しかし，段階的な値をそのまま利用して重要度付きの顕著性物体検出における真値としての利用も可能である。このデータセットは，はじめに [55] によって 850 枚の画像に対して物体領域に沿ったマスクを作成している。そしてそれらの画像に対し，12 人の被験者が画像をクリックすることで物体が顕著であるかどうかを二値で選ぶ。その作業によっ

て、PASCAL-S データセットにおける顕著性物体検出の真値は図 18 のように $[0, 12]$ の整数値を与えられ $[0, 255]$ に正規化された後 png 形式で保存されている。すなわち、12 人の被験者すべてが選択した物体には 255 の重要度が与えられる。したがって、PASCAL-S データセットは 13 段階の重要度を持った重要度付き顕著性物体検出のデータセットであるといえる。

4.1.2 FCN

FCN はセマンティックセグメンテーションのために提案された CNN アーキテクチャの一つである [56]。これは、Visual Geometry Group (VGG) と呼ばれるネットワークを基に作られておりセマンティックセグメンテーションのみならず様々なアプリケーションに利用されている [35, 37, 38, 40]。VGG のアーキテクチャはそれぞれ 2 つあるいは 3 つの畳み込み層とプーリング層で構成された 5 つのブロックを持つ構造になっている。FCN は VGG アーキテクチャの最後の層を 1 チャンネルの畳み込み層に変更したものである。また、いくつかの手法では、さらに結合層や畳み込み層を加えることで優良な結果を得ている [35, 40]。これらの手法で結合層や畳み込み層の追加を行うのは、浅い層と深い層を結合することで大域的特徴と局所的特徴の両方を考慮したアーキテクチャを構築できるからである。



図 18 (a) 入力画像と (b) 重要度付き顕著性物体検出の真値

表 6 PASCAL-S データセット内の画像が含む物体数とその割合 [2]

# 顕著性物体	1	2	3	4	5	6	7+
# 画像数	300	227	136	72	43	28	44
割合 (%)	35	27	16	8	5	3	5

4.1.3 位置バイアス

顕著性検出と顕著性物体検出の多くの手法が物体の位置情報を事前情報として利用している [1, 12, 14, 16, 26, 36]. これは、通常人間が写真をとる時に被写体を画像の中心に置くため、顕著性物体が画像中央に位置しやすいからである. この性質を事前情報として利用するために、いくつかの顕著性物体検出手法が画像の中央付近に大きな重みをつけて検出を行う提案をした [14, 26]. また、それにならって顕著性検出手法でも [36] が FCN のアーキテクチャに位置バイアス畳み込み層を結合し、画像の中央付近に大きな重みを与える. これにより、顕著性物体の位置情報を事前情報とした高精度な検出を行っている.

4.1.4 順位付き顕著性物体検出

この手法は、顕著性物体とそれらの順位を推定する手法であり、本研究ではこの手法を RSOD と呼ぶ [27]。RSOD の CNN アーキテクチャは、粗い顕著性物体検出と詳細な顕著性物体検出を再帰的に計算して、これらの結果を最終的に結合する。このため、最終的な顕著性物体検出は粗い特徴と詳細な特徴両方を考慮できる。また、PASCAL-S データセットの真値を段階的に分解し学習データとすることで段階的な値のピクセル値を出力できる。追加の処理として、この手法は出力された顕著性マップからそれぞれの顕著性物体の順位を推定している [27]。基本的に、順位は物体内のピクセル顕著性値の合計を次のように平均することで計算される。

$$\text{Rank}(S(X)) = \frac{\sum_{i \in \Omega_X} \chi_i}{N_X}, \quad (18)$$

ここで、 S 、 X 、 Ω_X 、 χ_i および N_X はそれぞれ予測された顕著性マップ、顕著性物体の候補、 X に属するピクセルのインデックスのセット、 i 番目のピクセルの顕著性値、および X のピクセルの総数を示している。[27] で明確に説明されていないため、計算値が正規化されているかは不明である。この手法では、物体領域に PASCAL-S の真値の顕著性物体領域を利用している。実験では、この方法は従来の評価指標であるスピアマンの順位相関を単純に利用している [57]。結果のスコアは $[0, 1]$ で線形に正規化され、トレー

ニングで利用しなかった PASCAL-S の画像がそのままテスト画像に利用されている。

4.1.5 ピラミッドプーリング

ピラミッドプーリングは、セマンティックセグメンテーション手法である [41] で提案された構造である。特徴抽出のアーキテクチャの後に、異なるサイズとストライドを持ついくつかのプーリング層を並列に配置し、これらの層をアップサンプリングしたのちに結合する手法である。サイズとストライドの異なるプーリング層は、それぞれ異なる受容野を持ち、局所的な情報と大域的な情報の両方を考慮できる。これによって、この構造は画像内の物体のサイズの違いによる誤検出を防ぐことができる。本提案でも、物体のサイズに関わらない推定が必要であるため、ピラミッドプーリングを利用している。

4.1.6 インスタンスセグメンテーション

インスタンスセグメンテーションは、画像内から物体を抽出するだけでなく物体ごとに境界を識別する手法である。物体認識やセマンティックセグメンテーションよりも詳細に画像を分割する方法として近年、様々なインスタンスセグメンテーションが盛んに研究されている。中でも、[42] は高精度な推定が可能であるため、本手法ではこれを物体領域の推定に利用する。ただし、本手法の一段階目には様々な手法が利用できるため [43, 58, 59] などの手法で物体領域を推定することも可能である。

4.2 重要度について：顕著性物体の重要度推定

4.2.1 重要度の定義

本研究では、重要度付き顕著性物体検出は顕著性物体の輪郭と、それらの重要度を推定する手法であると定義する。この手法では、物体領域上のピクセル値が物体の重要度を表すように推定結果を生成する。重要度付き顕著性物体検出の出力は、顕著性物体検出の出力とほぼ同等であるが、顕著性物体検出の出力は二値の画像であるのに対して、重要度付き顕著性物体検出は図 18 に示すように物体に対して N 段階の値が与えられる。 N 段階は、出力が $[0, N - 1]$ の整数値を持つことを意味しており、0 は顕著性がないことを示している。4.1.1 節で説明した PASCAL-S と同様に、線形に正規化される。そのため、重要度付き顕著性物体検出は顕著性物体検出を一般化したものであり、従来の顕著性物体検出は $N = 2$ の重要度付き顕著性物体検出であるといえる。

本研究では、データセットを分析したことによって得た自然画像の特性に基づいて $N = 7$ 段階の重要度を使用する。表 6 は、PASCAL-S 内の顕著性物体の数を示している [2]。ここで、一行目と二行目および最後の行はそれぞれ顕著性物体の数、その画像の枚数、全体からの割合を示している。8 列目の 7+ は 7 つ以上の顕著性物体を示している。

表 6 より、自然画像には通常 6 つ以下の顕著性物体が含まれており、7 つ以上の顕著性物

体が含まれていることは稀であることがわかる。また，7つ以上の物体を含む画像のほとんどが集合写真や駐車場など，同一の重要度を持つ物体が複数含まれている画像である。したがって，7段階での重要度は自然画像の重要度を適切に表現できるといえ，一般的に有効であるといえる。ただし， N の値は様々な画像アプリケーションに対して柔軟に変更できるため，場合によってはより多数の重要度段階を持つ場合もある。

4.2.2 重要度の有効性

重要度付き顕著性物体検出は，顕著性物体検出を一般化した方法である。通常，人間は画像内の物体に対して自分の興味の度合いで重要度付けを行うことがわかっている。さらに，実験より被験者は画像内の物体の位置によって物体への重要度付けが変化することがわかっている。より人間の特性に忠実な推定を行うため，重要度付き顕著性物体検出は顕著性物体を検出するのに加え人間が行う重要度付けの結果を推定する。さらに，後処理として様々なパラメータを利用して閾値を設定することで重要度付き顕著性物体検出は，顕著性物体検出の結果として利用することもできる。重要度付き顕著性物体検出は，画像リターゲティング，コンテンツに応じた画像コーディング，画像要約などの様々な画像処理に利用できる。例えば，実験より重要度付き顕著性物体検出は顕著性物体検出よりも画像リターゲティングに適していることがわかる。図 19 は，複数の物体が写っている画像を [8] によってリターゲティングした結果を示している。図 19 の (a) の入力画像は犬が

人間の乗ったそりを引いているという意味を読み取ることができる。この文より、重要な単語は犬、人間、乗る、そり、引くとなる。画像リターゲティングでは、文章の意味を損なわず、入力画像の重要な単語や文を結果に保持しておく必要がある。画像の意味を保存することを考慮する場合、(d)は犬がはっきり見えないため、「何かが人間の乗ったそりを引いている」という文章になってしまい、リターゲティングの結果として不十分である。反対に、(e)では重要度付き顕著性物体検出の結果を利用したリターゲティング画像が「犬が人間の乗ったそりを引いている」という元の文章を正確に表現している。ほかの画像においても、重要度付き顕著性物体検出の結果は画像リターゲティングにおいて顕著性物体検出結果に対して優位性を示すことが多い。

4.2.3 評価指標

重要度付き顕著性物体の評価指標は、物体領域を正確に検出していることと重要度を正確に検出していることの二つを測定する必要がある。重要度付き顕著性物体検出も、顕著性物体検出と同様に物体の輪郭は正確に検出しなくてはならないため顕著性物体検出と同様の方法で評価する必要がある。さらに、重要度の値について真値と比較する必要がある。真値画像内において、別の物体より重要度が高い物体は重要度付き顕著性物体検出の結果の中でも高い重要度を示すはずであり、真値と重要度付き顕著性物体検出の値は近いほど優れているといえる。そのため、従来手法では二つの値の順位がどれほど類似してい



(a) 入力画像



(b) 顕著性物体マップ



(c) 重要度付き顕著性物体マップ



(d) (b) による結果



(e) (c) による結果

図 19 複数物体画像における顕著性物体検出と重要度付き顕著性物体検出を用いた画像リターゲティングの結果

るかを測定するスピアマンの順位相関係数を利用していた。しかし、従来の順位相関係数

を用いた場合、値の前後関係について相関を評価できるが実際の値の類似性について無視

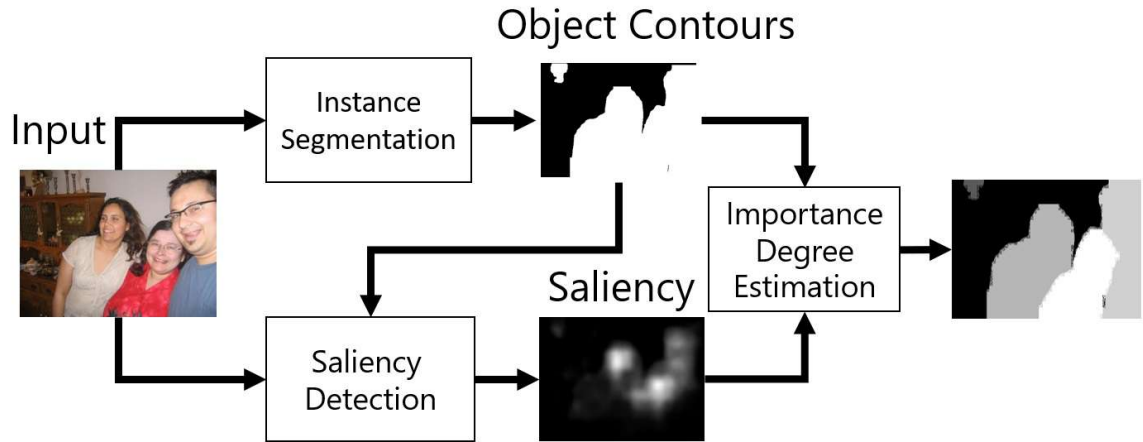


図 20 重要度付き顕著性物体検出の提案手法概略図

するため，[27] で利用されている順位相関係数は重要度の計算には適していない．本研究では，重要度付き顕著性物体検出の評価指標を提案する．はじめに，物体領域の正確さについては従来と同様に F-measure が利用できる．重要度付き顕著性物体検出の評価指標は，F-measure と後述する提案重要度評価指標との線形結合またはそれらを並列に使用して定義できる．提案された指標 F は，相関とスコアの類似性の評価指標を単純に組み合わせる以下のように定義される

$$F(\mathbf{v}_p, \mathbf{v}_t) = \alpha R(\mathbf{v}_p, \mathbf{v}_t) + (1 - \alpha)I(\mathbf{v}_p, \mathbf{v}_t) \quad (19)$$

ここで， R ， I ， α ， \mathbf{v}_p ，および \mathbf{v}_t は，相関係数および類似性の評価指標，バランスのためのフリーパラメータ，および各要素がそれぞれの物体の重要度であるベクトルをそれぞれ示している．提案法では，ケンドールの順位相関係数を R として使用する [60]．こ

表7 PASCAL-S データセットにおける重要度推定法別の評価値 [2]

	Sum.	Ave.
ケンドール	0.846	0.726
スピアマン	0.864	0.737

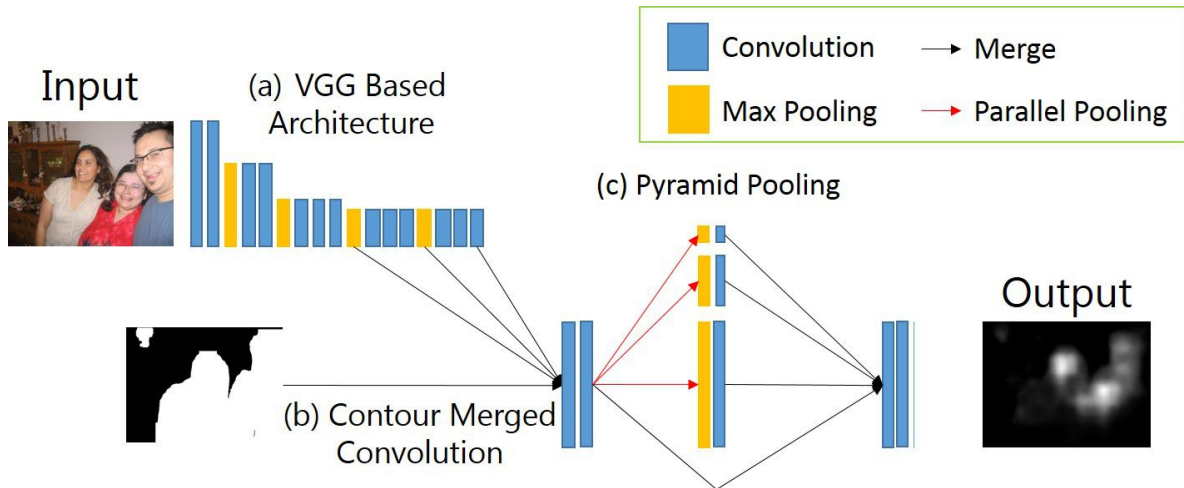


図21 提案 CNN のアーキテクチャ

これは、重要度の前後関係のみを直接評価するためスピアマンの順位相関係数よりも適して

いるからである。I は二乗誤差を使用して、次のように定義する。

$$I(\mathbf{v}_p, \mathbf{v}_t) = \frac{1}{N} \sum_{i=1}^N \exp(-(v_{pi} - v_{ti})^2 / (2\sigma^2)) \quad (20)$$

ここで、 N 、 v_{pi} 、および v_{ti} はそれぞれ物体の数、および \mathbf{v}_p と \mathbf{v}_t の i 番目の要素を示

し、 σ はガウス分布の分散を制御する自由パラメータである。 $[-1, 1]$ で実数値を出力す

る R は $[0, 1]$ に線形正規化され、 I は (20) より $[0, 1]$ での実数値を持つ。したがって F は

$[0, 1]$ の値を持つ。この評価指標の提案には多くの実験的証拠が必要であるが、 F の有効

性を実験の章に簡単に示し、これに関する詳細な研究は今後の作業として残す。

4.2.4 データセットの作成

重要度付き顕著性物体検出のデータセットを作成するには、4.1.1で説明した PASCAL-S の手順が適している。物体領域の真値は手作業によってマスキングすることで取得し、重要度は次のように決定する。はじめに、重要度推定のためには実験対象者に一つの画像内の顕著性物体に対し重要度を付加するよう求める必要がある。しかし、重要度の基準は個人によって異なるので、重要度付加のための厳格なルール付けは難しいといえる。そのため、PASCAL-S の方法では被験者に興味深い物体を指定し、顕著かどうかの二値で重要度付けするように求めるだけとしている。この実験の後、顕著であると認識した対象の数を直接重要度の値として物体に付加し、重要度を決定する。重要度は、各顕著性物体の物体領域に沿ってすべてのピクセルが均一に同じ値を持つ。以上より、 M 人の被験者がこの実験に適用された場合、この重要度は M 段階になる。この方法はシンプルだが、一般的なデータセットの作成には多数の被験者が必要になる欠点がある。被験者を使った実験避けるため、既存の顕著性検出データに基づく重要度付き顕著性物体検出の準備手順を示す。前述のように、主観的な実験には多くの人と多額の費用がかかる問題がある。これを回避するために、既存の顕著性検出データセットを使用して重要度付き顕著性物体検出の真値を作成する。本研究では、顕著性検出の真値から得られたピクセル値を顕著性物体上

に合計し重要度の値としてみなす。これは、一つの画像上で次のように定義される。

$$Deg_i = \frac{\sum_{j \in \Omega_i} s_j}{\max_i \{ \sum_{j \in \Omega_i} s_j \}} \quad (21)$$

ここで、 Deg_i 、 s_j 、および Ω_i はそれぞれ、 i 番目の物体の重要度、顕著性マップの i 番目のピクセル値、および i 番目の物体内のピクセルのインデックスのセットを示している。重要度付き顕著性物体のマップを作製するためには、 i 番目の物体内のピクセル値を Deg_i として均一に設定し、結果のマップを N を使用して線形に量子化する。顕著性物体検出の真値は各ピクセルの顕著性の度合いを示すため、合計値は物体の重要度とほぼ同等であるといえる。同様に、顕著性検出の真値内の物体のピクセル値は、物体を認識して顕著として分類する対象の数とほぼみなされるため対象の数が多い場合は合計を計算する。これは 4.1.1 節で述べた PASCAL-S における方法と同様である。上記の過程に基づいて提案された手順は重要度付き顕著性物体検出データセットの作成に有効である。上記の手順は、4.1.4 節で言及された RSOD 手順と比較して高い妥当性を持っていることを実験的に示す [27]。これらの手順を使用して、重要度付き顕著性物体検出真値を PASCAL-S の視点データと物体領域マスクにより作成する。表 7 は、この比較を示している。ここで Ave. と Sum. はそれぞれ平均と合計を示し、それぞれ RSOD による方法と提案による方法を示している。つまり、PASCAL-S の顕著性物体検出真値と視点データを使用した推定値との間の評価を示している。簡単にするために、スピアマンとケンドール両方の順位

相関係数を評価指標として使用する [57,60]. 表7より, 提案された手順は **RSOD** での手順よりも明らかに優れているため, 上記の意見が有効であることが示された

4.3 提案法：顕著性物体の重要度推定

4.3.1 概要

提案した重要度付き顕著性物体検出手法の概要について図20に示す. この手法は, インスタンスセグメンテーション, 顕著性検出および重要度の推定の3つのブロックで構成されている. まず, 入力画像にインスタンスセグメンテーションを適用し, 物体領域を抽出する. ここでは, [43,58,59]などの任意の手法を物体領域の抽出に利用できる. 次に, 入力画像に対して最初のブロックで抽出された物体領域を利用して, 提案されたCNN手法によって顕著性検出を行う. 最後に, 提案法は第1ブロックと第2ブロックの結果から重要度推定を行い, N 段階の重要度付き顕著性物体検出結果を出力する. それぞれのブロックは独立しているため, この手法は多様な拡張性があり重要度付き顕著性物体の基本設計として機能する.

4.3.2 提案顕著性

概要

本節では、第2ブロックである顕著性検出について説明する。提案顕著性検出では、入力画像と最初のブロックで検出された物体領域をCNNで学習する。提案法のアーキテクチャは、入力の一部として物体領域を利用しそれらの多重解像度の特徴を抽出して顕著性検出を行う。損失関数は、物体領域に基づいて背景領域と物体領域に異なる重みを適用する。提案されたCNNは従来の顕著性検出及び顕著性物体検出の位置バイアスと同様の方法で物体領域を考慮している。

アーキテクチャ

図21と表8は、それぞれ提案されたCNNのアーキテクチャとそのパラメータを示している。図21(a) - (c)はそれぞれ表の(a) - (c)に対応している。表でConv., Pool., およびp*と書かれている部分は、それぞれ畳み込み層、マックスプーリング、およびピラミッドプーリングを示している。畳み込み層の活性化関数にはReLU (rectified linear unit) を利用している [61]。VGGベースの手法によって、図21(a)により入力画像の特徴を算出する。さらに、第1ブロックで検出された物体領域は、Pool. 3, Pool. 4 および Conv. 5-3 はチャンネル方向に結合され、結合したものは Conv. 6-1 に入力される。その

後, [41] で提案されているピラミッドプーリングによって大域的情報と局所的情報を効果的に取得し, 最後に1チャンネルの顕著性検出結果として出力される.

損失関数

提案されたCNNの損失関数は, 顕著性領域と物体領域にそれぞれ大きな重みと中程度の重みを適用する. 背景領域は, 顕著性物体検出にとって興味領域ではないため, 低い重みを適用する. これより, 損失関数 L は以下のように定式化される

$$L(\mathbf{w}) = \frac{1}{N} \sum_{i=1}^N \left\| \frac{\frac{\phi(\mathbf{x}_i)}{\max \phi(\mathbf{x}_i)} - \mathbf{y}_i}{\beta - (\mathbf{O}_i + \mathbf{y}_i)} \right\| \quad (22)$$

ここで \mathbf{y}_i , \mathbf{x}_i , \mathbf{O}_i , $\phi(\cdot)$, β はそれぞれ顕著性の真値, 顕著性の推定値, 物体領域のマスク, 正規化関数, およびフリーパラメータを示している. マスクは, インスタンスセグメンテーションの結果を二値化して生成している. $\phi(\cdot)$ は顕著性の推定値を $[0, 1]$ に正規化する. 通常, β を2または $\mathbf{O}_i + \mathbf{y}_i$ の最大値に設定する. i 番目のピクセルが顕著性を持ちかつ物体内にある場合, $\beta - (\mathbf{O}_i + \mathbf{y}_i)$ の値は低くなるため, そのピクセルには高い重みが割り当てられることになる.

モデルのトレーニング

モデルのトレーニングには, 4.3.2 節の損失関数と, COCO および SALICON のデータセットを利用した [3,4]. COCO データセットには自然画像とその物体領域マスクが含ま

れており，SALICONには同様の画像に対して顕著性マップが含まれている．顕著性マップは $\tau = 0.15$ の閾値を用いて二値化され，物体領域外の顕著性に関しては0として設定した．最適化には確率的勾配降下法を使用し，Nesterov momentum，減衰率，学習率をそれぞれ0.9，0.5，および 10^{-3} に設定している [62]．損失関数の β は2.3に設定した．これは，顕著性領域，物体領域，および背景領域の比率から経験的に決定している．

4.3.3 重要度推定

本研究では，重要度推定は4.2.4節で提案した手順と同様にして行う．物体領域は第一ブロックで既に検出されており，それらに対する顕著性も第2ブロックで検出されているため，3番目のブロックではこれらを利用して重要度の推定を行う．このブロックでは，物体領域上にあるすべてのピクセルの顕著性を合計し，その結果が重要度のスコアになる．最後に，4.2.4節で提案された手順と同様に，重要度付き顕著性物体検出結果は N で線形に量子化される．

4.4 シミュレーション：顕著性物体の重要度推定

4.4.1 実験概要

本節では、顕著性物体検出と重要度付き顕著性物体検出の提案法と最新手法を比較する。はじめに、4.4.2節では4.2.3節で提案された評価指標の有効性について説明する。次に、4.4.3節では顕著性物体の推定精度について顕著性物体検出の最新手法と比較を行う。最後に、4.4.4節で重要度推定精度について最新の順位付き顕著性物体検出手法との比較を行う。本研究では、第一ブロックに [42] で提案されたインスタンスセグメンテーションを利用している。これは、最新の手法ではないが一般的に精度よく物体領域を検出できる手法である。また、4.2.4節に基づいて COCO と SALICON のテストセットからデータセットを作成した。これらのデータセットにはそれぞれ物体領域のマスクとその視点情報が含まれている。本節では、提案手法は「提案法」と表記されている。

4.4.2 提案評価指標の妥当性

4.2.3節で述べた提案評価指標の有効性を本節で簡単に示す。表9は、任意のベクトルに対するスピアマンとケンドールの順位相関係数の数値を示している。表9の任意ベクトルのペアは上からそれぞれ以下の状態を示している。上から、順位は同等であるが実際の

値がわずかに異なっているもの，実際の値は同等だが順位がわずかに違うもの，順位は同等だが実際の値が大幅に異なっているもの，値は同等だが順位が大幅に違うものの4つである。4.2.3節で述べたように，顕著性物体検出の評価指標は順位相関と実値の類似性を同時に評価する必要がある。一番目と三番目の結果では，従来の評価指標と提案評価指標で大きな差はなく，妥当な評価であるといえる。しかし，二番目と四番目では順位相関係数は順位の差によって大きく値が変動し，重要度付き顕著性物体検出の評価指標としては適していないことがわかる。特に四番目のペアでは，順位相関の評価では非常に低い数値であるが，実際の値はほぼ同じであるため，物体の重要度もほぼ同等であるとみなされるべきである。そのため，順位相関係数での評価より，提案評価指標による評価の方が，より利用先のニーズに適した評価指標であるといえる。

4.4.3 顕著性物体検出としての比較

設定

本実験では，HDCT [16]，RFCN [17]，DHS [22]，DSSOD [26]，RSOD [27] を比較対象とした。これらの手法を DUTS，PASCAL-S，SALICON ベースのデータセットによって比較した [2-5]。評価指標には F-measure を用い，段階的な重要度を持つ PASCAL-S データセットと SALICON ベースのデータセットには 0 領域を非顕著，1 以上の値を持つ領域を顕著として二値化した [48]。また，それぞれの手法の検出結果も HDCT，RFCN，

DHS, DSSOD の結果は閾値 0.14 で二値化され, RSOD と提案法の場合は重要度 1 を閾値として二値化した。

実験結果

表 10, 11, および 12 は, 各データセットでの各手法の F-measure での評価値を示している。図 24 および 25 は入力画像, それらの真値, および閾値処理前のそれぞれの手法の結果を示している。ここで, 平均は各データセット内のすべての画像の平均値を示している。PASCAL-S および SALICO ベースデータセットの画像には通常, 複数の物体が含まれており, 対照的に DUTS データセットには通常一つの物体が大きく写っている。表 10 では, 提案法は DUTS データセットの数値結果が従来法より劣っている。しかし, 表 11 と 12 では他の手法を上回る結果を示し, 図 24 と 25 でもその様子がわかる。表 11 と 12 では提案法が従来法より正確に物体領域を推定していることを示している。特に提案法では図 25 の駐車場画像とパーティー画像において不正確な推定を抑制している。しかし, 表 10 の上半分に示すようにインスタンスセグメンテーションの方法は上記特性のために DUTS に対して何も検出しないことが多かった。提案の結果は, その場合を除いて他の方法の結果と概ね同等であった。提案法は物体領域を正確に検出するインスタンスセグメンテーションを用いれば, この問題に対応することができる。

4.4.4 重要度推定としての比較

設定

重要度付き顕著性物体は，他に研究されている手法がないため順位付き顕著性物体検出の手法である RSOD と比較した．データセットは PASCAL-S と SALICON ベースのデータセットを使用し，順位相関と提案評価指標によって評価を行った．ここで， α および σ はそれぞれ経験的に 0.5 および 2.3 に設定した．実験では，それぞれの結果とデータセットの真値は $N = 7$ で均一に正規化されている． α および σ はフリーパラメータであるため，ユーザー設定である N に合わせて任意の値をとることができる．しかし，図 22 と図 23 に示すように提案法の優位性はこれらのパラメータによって大きく変動せず，従来法に対して提案法は有意な結果を示し続ける．そのため，本実験では経験的に α および σ の値を設定したが， N に合わせてこれらの値を変更しても，提案法の優位性は変化しない．

実験結果

表 13 と表 14 は，各データセットの各評価指標における RSOD と提案法のスコアを示す．図 24 と図 25 はデータセット内の画像とその真値，およびそれぞれの手法の結果を示している．表 13 と表 14 から，提案手法は RSOD を上回っていることがわかる．図 24 および図 25 より，提案法は物体の重要度を正確に推定していることがわかる．特に，パー

ティー画像，女性画像，男性画像では，高い顕著性を持つ物体に囲まれた小さな物体に低い顕著性を正しく推定推定できていることがわかる。

4.4.5 第4章のまとめ

第4章では，従来考慮されていなかった顕著性物体検出の重要度推定について概説し，重要度推定が画像リターゲットングなどのアプリケーションに拡張性をもたらし技術であることを述べた。4.1節では，顕著性物体検出でよく用いられる PASCAL-S データセット，FCN，位置バイアス，ピラミッドプーリングについて述べた。また，本研究と関連するインスタンスセグメンテーションや順位付き顕著性物体検出について述べ，その差異について述べた。4.2節では，従来考慮されていなかった重要度についての定義とその有効性について詳細な議論を行った。また，従来利用されていた評価指標について調査し，より重要度推定に適した評価指標の提案を行った。4.3節では，重要度付き顕著性物体検出を提案した。重要度は，視点データと物体領域の両方を学習した機械学習モデルによって推定され，推定された重要度を物体ごとに平均化することで各顕著性物体の重要度を推定した。4.4節では，提案された重要度付き顕著性物体検出と近い目的を持つ RSOD を比較対象として重要度の推定精度比較を行った。比較には，従来使用されていた順位相関と本論文で提案した評価指標を用いている。また，顕著性物体の検出精度についても従来の5手法と比較している。検出精度の実験では，単純なデータセットにおいては提案法はやや

低い精度であったが，複雑なデータセットになるにつれて誤検出が低減され有効性を示した．重要度推定の実験では，従来法に対して順位の精度と実値での精度両方において定量的かつ知覚的に優れていることを示した．

表8 提案 CNN アーキテクチャの構造

	Name	Size	Stride	Channel
(a)	Conv. 1-1	3×3	1	64
	Conv. 1-2	3×3	1	64
	Pool. 1	2×2	2	64
	Conv. 2-1	3×3	1	128
	Conv. 2-2	3×3	1	128
	Pool. 2	2×2	2	128
	Conv. 3-1	3×3	1	256
	Conv. 3-2	3×3	1	256
	Conv. 3-3	3×3	1	256
	Pool. 3	2×2	2	256
	Conv. 4-1	3×3	1	512
	Conv. 4-2	3×3	1	512
	Conv. 4-3	3×3	1	512
	Pool. 4	2×2	1	512
	Conv. 5-1	3×3	1	512
	Conv. 5-2	3×3	1	512
Conv. 5-3	3×3	1	512	
(b)	Conv. 6-1	3×3	1	512
	Conv. 6-2	3×3	1	512
(c)	pPool. 1	2×2	2	512
	pConv. 1	3×3	1	64
	pPool. 2	2×2	4	512
	pConv. 2	3×3	1	64
	pPool. 3	2×2	5	512
	pConv. 3	3×3	1	64
	Conv. 7-1	3×3	1	512
	Conv. 7-2	3×3	1	64
	Conv. 7-3	3×3	1	1

表9 任意のベクトルに対する順位相関と提案評価指標のスコア

ベクトル		スピアマン [57]	ケンドール [60]	提案評価指標
[2,6,3]	[1,5,3]	1.000	1.000	0.961
[4,1,6]	[4,2,4]	0.834	0.667	0.748
[6,2,7]	[2,1,3]	1.000	1.000	0.692
[2,1,3]	[1,3,2]	0.250	0.333	0.562

表10 DUT データセットにおける F-measure [5]

画像	HDCT [16]	RFCN [17]	DHS [22]	DSSOD [26]	RSOD [27]	提案法
画像 1	0.120	0.253	0.745	0.608	0.000	0.000
画像 2	0.396	0.330	0.945	0.933	0.000	0.000
画像 3	0.732	0.794	0.750	0.748	0.875	0.875
画像 4	0.669	0.614	0.894	0.898	0.904	0.904
Average	0.521	0.509	0.761	0.746	0.711	0.686

表11 PASCAL-S データセットにおける F-measure [2]

画像	HDCT [16]	RFCN [17]	DHS [22]	DSSOD [26]	RSOD [27]	提案法
バイク	0.875	0.704	0.879	0.911	0.840	0.840
車	0.284	0.283	0.597	0.919	0.960	0.960
犬	0.575	0.728	0.749	0.917	0.940	0.940
馬	0.885	0.699	0.928	0.898	0.928	0.928
交通	0.687	0.688	0.772	0.935	0.965	0.965
平均	0.623	0.584	0.778	0.788	0.793	0.794

表 12 SALICON ベースのデータセットにおける F-measure [3,4].

画像	HDCT [16]	RFCN [17]	DHS [22]	DSSOD [26]	RSOD [27]	提案法
駐車場	0.687	0.501	0.766	0.736	0.755	0.781
パーティ	0.565	0.684	0.817	0.736	0.902	0.937
女性	0.696	0.719	0.876	0.731	0.936	0.936
靴	0.702	0.753	0.839	0.876	0.944	0.944
男性	0.528	0.600	0.495	0.372	0.419	0.528
野球	0.784	0.654	0.842	0.873	0.776	0.941
平均	0.449	0.429	0.526	0.581	0.708	0.727

表 13 PASCAL-S データセットにおける重要度推定のスコア [2]

画像	スピアマン [57]		ケンドール [60]		提案評価指標	
	RSOD [27]	提案法	RSOD [27]	提案法	RSOD [27]	提案法
バイク	0.869	0.936	0.908	0.956	0.925	0.948
車	0.146	1.000	0.000	1.000	0.471	0.902
犬	0.146	1.000	0.000	1.000	0.261	1.000
馬	0.000	0.834	0.000	0.908	0.321	0.822
道路	0.380	1.000	0.000	1.000	0.402	0.971
平均	0.372	0.419	0.300	0.327	0.457	0.467

表 14 SALICON ベースのデータセットにおける重要度推定のスコア [3,4]

画像	Spearman's [57]		Kendall [60]		Prop. metric	
	RSOD [27]	提案法	RSOD [27]	提案法	RSOD [27]	提案法
駐車場	0.574	0.607	0.698	0.786	0.461	0.779
パーティー	0.394	0.555	0.691	0.741	0.430	0.698
女性	0.000	0.583	0.000	0.754	0.118	0.501
靴	0.028	0.815	0.000	0.887	0.128	0.865
男性	0.586	0.701	0.758	0.887	0.511	0.859
野球	0.868	0.901	0.900	0.947	0.793	0.924
Average	0.321	0.562	0.434	0.490	0.478	0.507

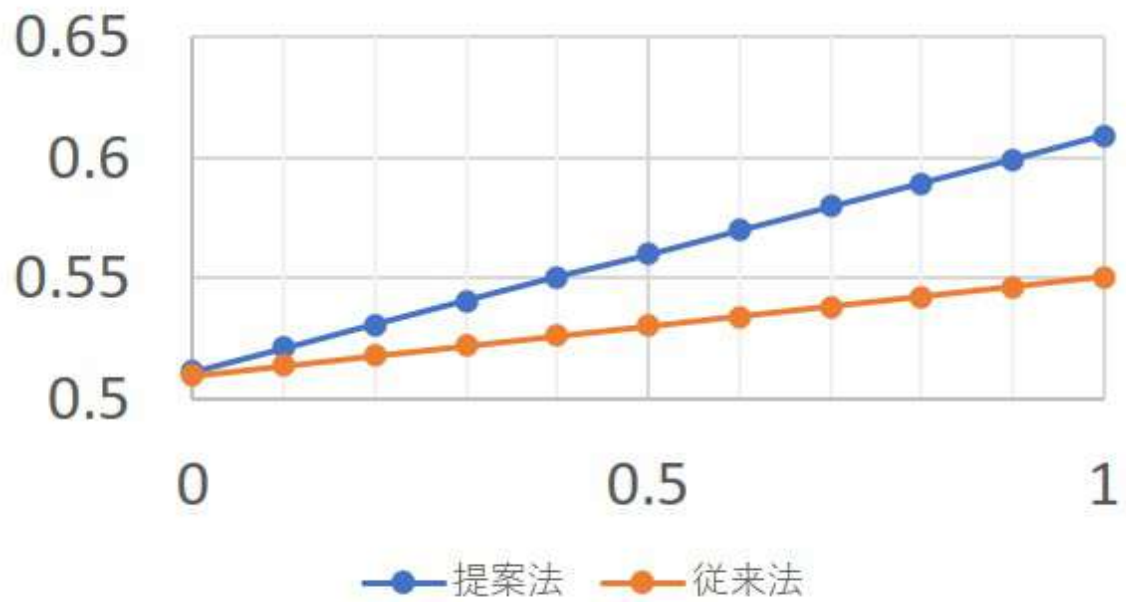


図 22 α に対する提案評価指標の変化：横軸は α 縦軸は評価値を表す

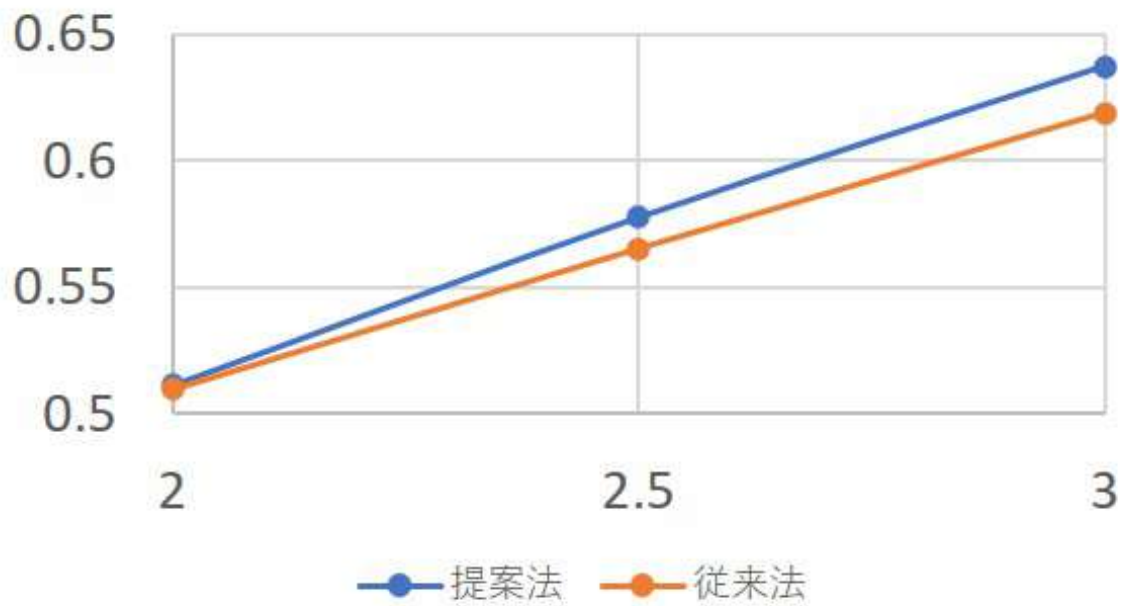


図 23 σ に対する提案評価指標の変化：横軸は σ 縦軸は評価値を表す



図 24 PASCAL-S データセットにおける結果画像 [2]

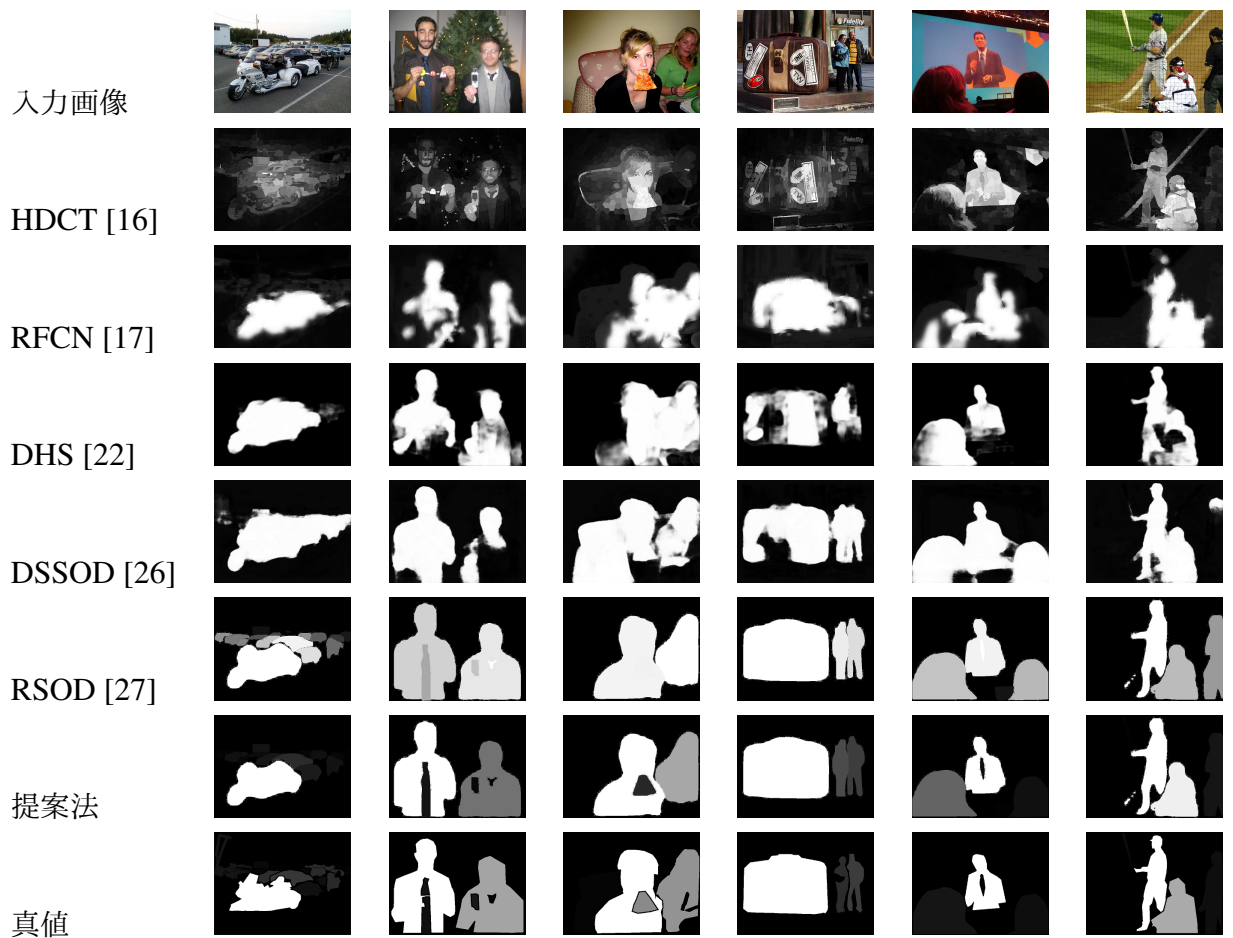


図 25 SALICON ベースのデータセットにおける結果画像 [3,4]

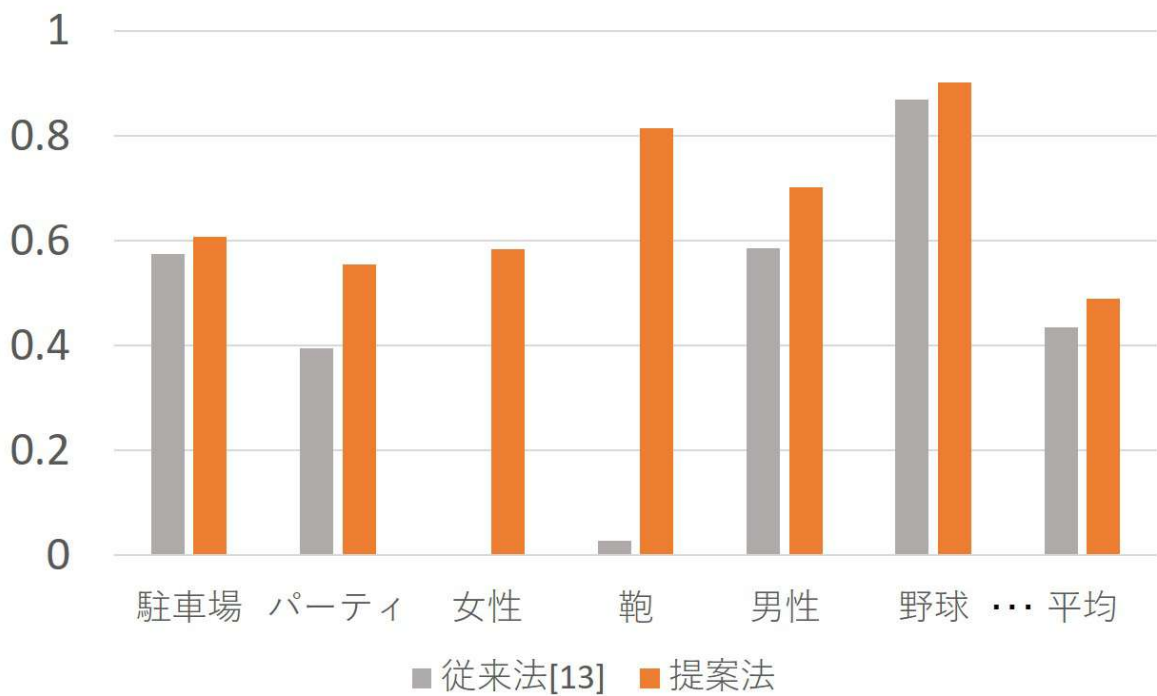


図 26 重要度推定比較:スピアマン相関係数

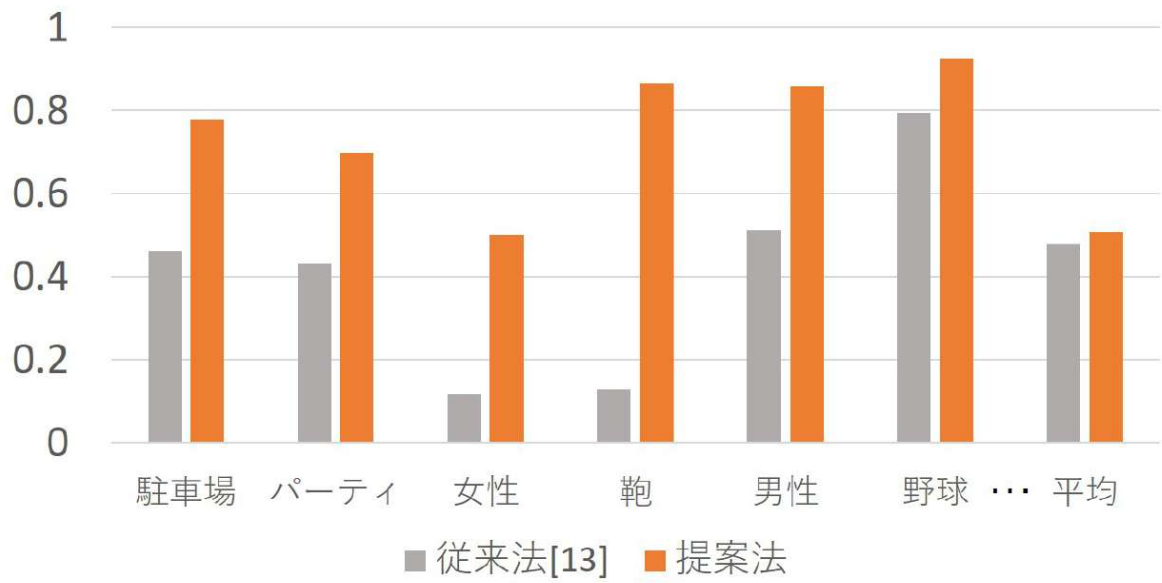


図 27 重要度推定比較:提案評価指標

第 5 章

結論

本論文では，顕著性物体検出の高精度化手法と顕著性物体の重要度推定手法を提案した．顕著性物体の高精度化手法では，従来手法では初期顕著性が不正確であること，加重結合によって誤検出が残存する問題があった．提案法では，正確な初期顕著性を提案することと競合型伝播を提案することで従来法の課題を解決した．初期顕著性の提案では，SIFT 特徴点と機械学習の利用により画像全体にまばらに配置され，かつ正確な初期顕著性を推定できた．また，信頼できる初期顕著性を伝播の最中に増加させることでより画像全体に均一な初期顕著性を実現している．実験では，従来法に比べて提案法は顕著性物体領域の推定精度において，定量的かつ知覚的に優れていることが示された．

顕著性物体の重要度推定手法では，従来非顕著あるいは顕著の二値であった推定を多値の推定を可能にした．また，重要度の定義や評価指標について詳細な議論を行った．評価

指標の提案では、従来では順位相関のみによる評価であったものを、提案では順位相関と実値での差を考慮した評価指標とした。これにより、よりリターゲティングなどの利用先に向けた評価が可能になった。また、提案手法では物体領域を考慮した CNN アーキテクチャと Loss 関数を提案し、N 段階での離散的な重要度推定を可能にした。実験では、提案法が各データセットに対して重要度の推定精度で他手法に比べ良好な成果を挙げており、提案法は従来法に比べ定量的かつ知覚的に優れていることが示された。

従来の手法では、リターゲティングなどの応用のために顕著性物体の検出精度を向上させることを目的としていた。これは、顕著性物体が物体の輪郭通りに検出されていないとリターゲティングの際に物体の欠損が起きるからである。そのため、3 章では顕著性物体の検出精度に関する研究を行った。3 章での提案により、顕著性物体が正確に検出され利用先であるリターゲティングの精度も向上した。しかし、すべての顕著性物体が正確に検出されていてもユーザーが大きなリターゲティングを行った場合、顕著性物体間での重要度の大小が考慮されていないために物体の変形や欠損が生じユーザーの要求通りの結果が得られない問題があった。そのため、4 章では顕著性物体の重要度推定に関する研究を行った。3 章と 4 章の研究により、顕著性物体の検出精度と重要度の付加両面において提案法は有効な成果を示し、画像リターゲティングへの広範囲な応用を示した。

また、4 章では物体領域の抽出に関してインスタンスセグメンテーション手法を利用しているが、物体領域抽出を 3 章の顕著性物体検出へ置き換えてさらに高精度な重要度付き

顕著性物体検出を行うことも可能である。3章の実験より、前景と背景の初期顕著性を正確に検出しそれを競合型伝播によって画像全体に伝播することで誤検出のない顕著性物体検出が可能であることが分かった。4章における実験では、顕著性検出の結果を重要度の推定に利用しており、これは3章における初期顕著性と類似したものである。そのため、3章での伝播手法を応用することにより顕著性物体検出と重要度推定を同時に行うことも可能である。しかし、この方法は物体領域の境界について推定できないためそれぞれの物体上に等しく重要度を推定できない。そのため、本研究ではこれを将来的な課題としインスタンスセグメンテーションによる物体領域の抽出を行っている。

最後に、4.2.3節でも述べたが、提案評価指標の有効性を示すためには多くの実験的証拠が必要である。実験の中で、 α や σ が評価に大幅な影響を及ぼさないことは述べたが、それぞれの数値の組み合わせや N の値に対する妥当な値は検証できていない。そのため、これら評価に関する研究は今後の課題である。

謝辞

本研究をすすめるに当たり，日頃から暖かくご指導して下さいました，本学電気電子情報工学専攻 岩橋政宏 教授，同 原川良介 助教および電気通信大学 情報理工学研究科 吉田太一 助教に心から感謝致します．また，審査を通じて貴重なご助言をいただきました，本学電気電子情報工学専攻 坪根正 准教授，同 圓道知博 准教授，同 杉田泰則 准教授および同 南部功夫 准教授に深く感謝致します．そして，本研究を進めるに当たり数々の助言，助力を頂いた岩橋研究室の皆様および弓削商船高等専門学校の皆様には心より感謝致します．最後に，大学院までの長い間，心身にわたって学業に専念する環境を整えてくれた両親に，この場を借りて心から感謝致します．

令和3年1月29日

参考文献

- [1] H. Li, H. Lu, Z. Lin, X. Shen, and B. Price, “Inner and inter label propagation: Salient object detection in the wild,” *IEEE Trans. Image Process.*, vol. 24, no. 10, pp. 3176–3186, 2015.
- [2] Y. Li, X. Hou, C. Koch, J. M. Rehg, and A. L. Yuille, “The secrets of salient object segmentation,” in *Proc. IEEE Conf. Comput. Vis. Patt. Recognit.*, 2014, pp. 280–287.
- [3] T.-Y. Lin, M. Maire, S. Belongie, J. Hays, P. Perona, D. Ramanan, P. Dollár, and C. L. Zitnick, “Microsoft coco: Common objects in context,” in *Proc. European Conf. Comput. Vis.* Springer, 2014, pp. 740–755.
- [4] M. Jiang, S. Huang, J. Duan, and Q. Zhao, “Salicon: Saliency in context,” in *Proc. IEEE Conf. Comput. Vis. Patt. Recognit.*, 2015, pp. 1072–1080.
- [5] L. Wang, H. Lu, Y. Wang, M. Feng, D. Wang, B. Yin, and X. Ruan, “Learning to detect salient objects with image-level supervision,” in *Proc. IEEE Conf. Comput. Vis. Patt. Recognit.*, 2017, pp. 136–145.
- [6] L. Marchesotti, C. Cifarelli, and G. Csurka, “A framework for visual saliency detection with applications to image thumbnailing,” in *Proc. IEEE Int. Conf. Comput. Vis.*, 2009, pp. 2232–2239.
- [7] M. Rubinstein, D. Gutierrez, O. Sorkine, and A. Shamir, “A comparative study of image retargeting,” vol. 29, no. 6, pp. 160–9, 2010.
- [8] A. Mansfield, P. Gehler, L. V. Gool, and C. Rother, “Scene carving: Scene consistent image retargeting,” in *Proc. European Conf. Comput. Vis.* Springer, 2010, pp. 143–156.
- [9] A. Jose and I. Heisterklaus, “Bag of fisher vectors representation of images by saliency-based spatial partitioning,” in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, 2017, pp. 1762–1766.
- [10] F. Perazzi, P. Krahenbuhl, Y. Pritch, and A. Hornung, “Saliency filters: Contrast based filtering for salient region detection,” in *Proc. IEEE Conf. Comput. Vis. Patt. Recognit.*, 2012, pp. 733–740.

- [11] H. Jiang, J. Wang, Z. Yuan, Y. Wu, N. Zheng, and S. Li, "Salient object detection: A discriminative regional feature integration approach," in *Proc. IEEE Conf. Comput. Vis. Patt. Recognit.*, 2013, pp. 2083–2090.
- [12] J. Sun, H. Lu, and X. Liu, "Saliency region detection based on markov absorption probabilities," *IEEE Trans. Image Process.*, vol. 24, no. 5, pp. 1639–1649, 2015.
- [13] R. Zhao, W. Ouyang, H. Li, and X. Wang, "Saliency detection by multi-context deep learning," in *Proc. IEEE Conf. Comput. Vis. Patt. Recognit.*, 2015, pp. 1265–1274.
- [14] N. Tong, H. Lu, X. Ruan, and M. H. Yang, "Salient object detection via bootstrap learning," in *Proc. IEEE Conf. Comput. Vis. Patt. Recognit.*, 2015, pp. 1884–1892.
- [15] Y. Qin, H. Lu, Y. Xu, and H. Wang, "Saliency detection via cellular automata," in *Proc. IEEE Conf. Comput. Vis. Patt. Recognit.*, 2015, pp. 110–119.
- [16] J. Kim, D. Han, Y.-W. Tai, and J. Kim, "Salient region detection via high-dimensional color transform and local spatial support," *IEEE Trans. Image Process.*, vol. 25, no. 1, pp. 9–23, 2016.
- [17] L. Wang, L. Wang, H. Lu, P. Zhang, and X. Ruan, "Saliency detection with recurrent fully convolutional networks," in *Proc. European Conf. Comput. Vis.* Springer, 2016, pp. 825–841.
- [18] T. Wang, L. Zhang, H. Lu, C. Sun, and J. Qi, "Kernelized subspace ranking for saliency detection," in *Proc. European Conf. Comput. Vis.*, 2016, pp. 450–466.
- [19] L. Zhang, C. Yang, H. Lu, X. Ruan, and M.-H. Yang, "Ranking saliency," *IEEE Trans. Patt. Anal. Mach. Intel.*, vol. 39, no. 9, pp. 1892–1904, 2016.
- [20] C. Sheth and R. V. Babu, "Object saliency using a background prior," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, 2016, pp. 1931–1935.
- [21] F. Yang and M.-H. Yang, "Top-down visual saliency via joint crf and dictionary learning," *IEEE Trans. Patt. Anal. Mach. Int.*, 2016.
- [22] N. Liu and H. J., "Dhsnet: Deep hierarchical saliency network for salient object detection," in *Proc. IEEE Conf. Comput. Vis. Patt. Recognit.*, 2016, pp. 678–686.
- [23] L. Zhang, J. Ai, B. Jiang, H. Lu, and X. Li, "Saliency detection via absorbing markov chain with learnt transition probability," *IEEE Trans. Image Process.*, vol. 27, no. 2, pp. 987–998, 2017.
- [24] G. Li, Y. Xie, L. Lin, and Y. Yu, "Instance-level salient object segmentation," in *Proc. IEEE Conf. Comput. Vis. Patt. Recognit.*, 2017, pp. 2386–2395.
- [25] J. Yang and M. H. Yang, "Top-down visual saliency via joint crf and dictionary learning," *IEEE Trans. Patt. Anal. Mach. Intel.*, vol. 39, no. 3, pp. 576–588, 2017.
- [26] Q. Hou, M.-M. Cheng, X. Hu, A. Borji, Z. Tu, and P. H. Torr, "Deeply supervised

- salient object detection with short connections,” in *Proc. IEEE Conf. Comput. Vis. Patt. Recognit.*, 2017, pp. 3203–3212.
- [27] M. Amirul Islam, M. Kalash, and N. D. B. Bruce, “Revisiting salient object detection: Simultaneous detection, ranking, and subitizing of multiple salient objects,” in *Proc. IEEE Conf. Comput. Vis. Patt. Recognit.*, 2018, pp. 7142–7150.
- [28] C. Aytekin, A. Iosifidis, and M. Gabbouj, “Probabilistic saliency estimation,” *Patt. Recognit.*, vol. 74, pp. 359–372, 2018.
- [29] R. Fan, M.-M. Cheng, Q. Hou, T.-J. Mu, J. Wang, and S.-M. Hu, “S4net: Single stage salient-instance segmentation,” in *Proc. IEEE Conf. Comput. Vis. Patt. Recognit.*, 2019, pp. 6103–6112.
- [30] L. Itti, C. Koch, and E. Niebur, “A model of saliency-based visual attention for rapid scene analysis,” *IEEE Trans. Patt. Anal. Mach. Int.*, vol. 20, no. 11, pp. 1254–1259, 1998.
- [31] J. Harel, C. Koch, and P. Perona, “Graph-based visual saliency,” in *Proc. Neural Information Process. Syst.*, 2006, pp. 545–552.
- [32] X. Hou and L. Zhang, “Saliency detection: A spectral residual approach,” in *Proc. IEEE Conf. Comput. Vis. Patt. Recognit.*, 2007, pp. 1–8.
- [33] R. Achanta, S. Hemami, F. Estrada, and S. Susstrunk, “Frequency-tuned salient region detection,” in *Proc. IEEE Conf. Comput. Vis. Patt. Recognit.*, 2009, pp. 1597–1604.
- [34] J. Zhang and S. Sclaroff, “Exploiting surroundedness for saliency detection: A boolean map approach,” *IEEE Trans. Patt. Anal. Mach. Int.*, vol. 38, no. 5, pp. 889–902, 2016.
- [35] M. Cornia, L. Baraldi, G. Serra, and R. Cucchiara, “A deep multi-level network for saliency prediction,” in *Proc. IEEE Int. Conf. Patt. Recognit.* IEEE, 2016, pp. 3488–3493.
- [36] S. S. S. Kruthiventi, K. Ayush, and R. V. Babu, “Deepfix: A fully convolutional neural network for predicting human eye fixations,” *IEEE Trans. Image Process.*, vol. 26, no. 9, pp. 4446–4456, 2017.
- [37] N. İmamoğlu, C. Zhang, W. Shmoda, Y. Fang, and B. Shi, “Saliency detection by forward and backward cues in deep-cnn,” in *Proc. IEEE Int. Conf. on Image Process.*, 2017, pp. 430–434.
- [38] R. Monroy, S. Lutz, T. Chalasani, and A. Smolic, “Salnet360: Saliency maps for omnidirectional images with cnn,” *Elsevier Trans. Signal Process. Image Communication*, vol. 69, pp. 26–34, 2018.
- [39] B. Hariharan, P. Arbeláez, L. Bourdev, S. Maji, and J. Malik, “Semantic contours from inverse detectors,” in *Proc. IEEE Int. Conf. Comput. Vis.*, 2011, pp. 991–998.

- [40] J. Long, E. Shelhamer, and T. Darrell, “Fully convolutional networks for semantic segmentation,” in *Proc. IEEE Conf. Comput. Vis. Patt. Recognit.*, 2015, pp. 3431–3440.
- [41] H. Zhao, J. Shi, X. Qi, X. Wang, and J. Jia, “Pyramid scene parsing network,” in *Proc. IEEE Conf. Comput. Vis. Patt. Recognit.*, 2017, pp. 2881–2890.
- [42] B. Romera-Paredes and P. H. S. Torr, “Recurrent instance segmentation,” in *Proc. European Conf. Comput. Vis.* Springer, 2016, pp. 312–329.
- [43] Y. Li, H. Qi, J. Dai, X. Ji, and Y. Wei, “Fully convolutional instance-aware semantic segmentation,” in *Proc. IEEE Conf. Comput. Vis. Patt. Recognit.*, 2017, pp. 2359–2367.
- [44] N. Dalal and B. Triggs, “Histograms of oriented gradients for human detection,” in *Proc. IEEE Conf. Comput. Vis. Patt. Recognit.*, vol. 1, 2005, pp. 886–893.
- [45] B. Su, S. Lu, and C. L. Tan, “Blurred image region detection and classification,” in *Proc. ACM Int. Conf. Multimed.*, 2011, pp. 1397–1400.
- [46] D. G. Lowe, “Object recognition from local scale-invariant features,” in *Proc. IEEE Int. Conf. Comput. Vis.*, vol. 2, 1999, pp. 1150–1157.
- [47] ———, “Distinctive image features from scale-invariant keypoints,” *Int. J. of Comput. Vis.*, vol. 60, no. 2, pp. 91–110, 2004.
- [48] R. Szeliski, *Computer Vision: Algorithms and Applications*. Springer London, 2010.
- [49] B. Alexe, T. Deselaers, and V. Ferrari, “Measuring the objectness of image windows,” *IEEE Trans. Patt. Anal. Mach. Int.*, vol. 34, no. 11, pp. 2189–2202, 2012.
- [50] L. Itti, “Automatic foveation for video compression using a neurobiological model of visual attention,” *IEEE Trans. Image Process.*, vol. 13, no. 10, pp. 1304–1318, 2004.
- [51] L. Breiman, “Random forests,” *Mach. Learn.*, vol. 45, no. 1, pp. 5–32, 2001.
- [52] T. Liu, Z. Yuan, J. Sun, J. Wang, N. Zheng, X. Tang, and H. Y. Shum, “Learning to detect a salient object,” *IEEE Trans. Patt. Anal. Mach. Int.*, vol. 33, no. 2, pp. 353–367, 2011.
- [53] H. Andrews and C. Patterson, “Singular value decompositions and digital image processing,” *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 24, no. 1, pp. 26–53, 1976.
- [54] T. Judd, F. Durand, and A. Torralba, “A benchmark of computational models of saliency to predict human fixations,” in *Proc. MIT Technical Reports*, 2012.
- [55] R. Mottaghi, X. Chen, X. Liu, N.-G. Cho, S.-W. Lee, S. Fidler, R. Urtasun, and A. Yuille, “The role of context for object detection and semantic segmentation in the wild,” in *Proc. IEEE Conf. Comput. Vis. Patt. Recognit.*, 2014, pp. 891–898.
- [56] K. Simonyan and A. Zisserman, “Very deep convolutional networks for large-scale image recognition,” in *Proc. Int. Conf. Learn. Representations*, 2015.
- [57] C. Spearman, “The proof and measurement of association between two things.” 1961.

- [58] K. Li, B. Hariharan, and J. Malik, “Iterative instance segmentation,” in *Proc. IEEE Conf. Comput. Vis. Patt. Recognit.*, 2016, pp. 3659–3667.
- [59] A. Arnab and P. H. Torr, “Pixelwise instance segmentation with a dynamically instantiated network,” in *Proc. IEEE Conf. Comput. Vis. Patt. Recognit.*, 2017, pp. 441–450.
- [60] M. G. Kendall, “A new measure of rank correlation,” *Biometrika*, vol. 30, no. 1/2, pp. 81–93, 1938.
- [61] V. Nair and G. E. Hinton, “Rectified linear units improve restricted boltzmann machines,” in *Proc. International Conf. on Mach. learn.*, 2010, pp. 807–814.
- [62] C. M. Bishop, *Pattern Recognition and Machine Learning*. Springer, 2006.

研究業績一覧

論文誌論文（査読付き）

(1) Yo Umeki, Taichi Yoshida and Masahiro Iwahashi, “Co-Propagation with Distributed Seeds for Salient Object Detection”, IEICE Trans. Information and Systems, vol. E101-D, no. 6, pp. 1640-1627, 2018.

(2) Yo Umeki, Isana Funahashi, Taichi Yoshida and Masahiro Iwahashi, “Salient Object Detection with Importance Degree”, IEEE Access, vol. 8, pp. 147059-147069, 2020.

国際会議（査読付き）

(1) Yo Umeki, Taichi Yoshida and Masahiro Iwahashi, “Estimation method of initial labels for propagation-based saliency detection,” in Proc. Asia-Pacific Signal and Information Processing Association Annual Summit and Conference (APSIPA ASC), 2016, pp. 1–4.

国内学会

(1) 梅木陽, 吉田太一, 岩橋政宏, “競合型ラベル伝播を基にした顕著性検出,” 第31回信号処理シンポジウム, 2016, pp. 1–6.

(2) 梅木陽, 吉田太一, 岩橋政宏, “複数回検出による顕著性検出,” 第32回信号処理シンポジウム, 2017, pp. 459–464.

(3) 梅木陽, 吉田太一, 岩橋政宏, “画像判別を用いた顕著性物体検出,” 第33回信号処理シンポジウム, 2018, pp. 312–317.