

Random Label Selection によるマルチラベルに対応した セマンティックセグメンテーション

福田 考晃[†] 藤吉 弘亘^{††}

[†] ヤマハ発動機株式会社 〒438-0017 静岡県磐田市新貝 2500

^{††} 中部大学大学院工学研究科 〒487-8501 愛知県春日井市松本町 1200

E-mail: [†]fukutat@yamaha-motor.co.jp, ^{††}hf@cs.chubu.ac.jp

あらまし セマンティックセグメンテーションは、高精度化を主眼に置き研究が進められてきた。一方、実用化を考えると対象の多様化への対応が必要不可欠である。我々は、対象の多様化を達成するために、マルチラベルを考慮したセマンティックセグメンテーション法を提案する。提案手法は、Random Label Selection という新しい手法を Random Forests に用い、Semantic Texton Forests によるセマンティックセグメンテーション手法をマルチラベルに拡張する。この手法を用いることにより、マルチラベルを捉えるための有効な特徴表現を行うことが可能となる。提案手法の有効性を示すために、提案手法の評価実験を行った。精度評価実験において、マルチラベルカテゴリを含んだ9カテゴリにおいて、再現率約0.7を確認し、従来法との比較実験において、マルチラベルカテゴリだけでなく、シングルラベルカテゴリにおいても性能向上を示すことができた。

キーワード Random Label Selection, Semantic Texton Forests, セマンティックセグメンテーション, マルチラベル

1. はじめに

屋内外の風景写真などの画像からシーンを理解するには、色分割だけでなく、同一カテゴリの領域を認識して分割する必要がある。これをセマンティックセグメンテーションと呼ぶ[10]。セマンティックセグメンテーションでは、画像内に存在する複数の物体カテゴリ名だけでなく、物体の位置を獲得することができるという利点があるため、シーンの理解には重要な技術である。

セマンティックセグメンテーションを実現する手法は、画素レベルや小領域レベルで統計的手法によりテキストチャ解析、色解析、Bag-of-features などの特徴量を算出し、Conditional Random Field(以下 CRF) や Markov Random Field モデルなどのグラフィカルモデルを利用することが一般的である。[10][4][11][12][6][2][3][7]。しかし、これらの従来法では、分割した領域に対応するカテゴリが1つである。この場合、花畑が撮影された画像のように花と茎が混在している対象領域を、それぞれ花と茎のカテゴリとして分割することになり、非常に難しい問題となる。一方、領域に対応する複数カテゴリを許容すれば、花畑の領域を花と茎のマルチラベルを持つ領域とすることができ、この問題を解決することができる。

そこで本稿では、マルチラベルに対応したセマンティックセグメンテーション法を提案する。提案手法では、高精度なセマンティックセグメンテーションを実現可能な Semantic Texton Forests [10] をベース手法とする。しかし、従来の Semantic Texton Forests で用いられる Random Forests は単一のラベルを対象としているため、本研究では、Random Label Selection を提案し、マルチラ

ベルに対応する。また、後段のセグメンテーションのための識別器には、Label Powerset 法を用いたシンプルな学習手法を利用する。

本稿の構成は、2. 節にて関連研究と問題設定について、3. 節にて Semantic Texton Forests について、4. 節にて提案手法を持ちいたマルチラベルに対応したセマンティックセグメンテーションについて、5. 節にて実験、考察を述べ、6. 節にてまとめる。

2. 関連研究と問題設定

2.1 関連研究

セマンティックセグメンテーション手法は、一般的に画素レベルや小領域レベルからテキストチャ解析や、色解析により特徴量を抽出して識別を行い、さらに、後段処理を設けコンテキスト情報を利用する。このようなアプローチにおいて従来から、コンテキストの利用法、特徴量の検討、学習サンプルの簡易化の3つの問題に分けて取り組まれている。コンテキストの利用については、主に CRF が用いられている。He 等は階層構造を取り入れた CRF を用いた手法を提案した [4]。Shotton 等は、CRF により捉える対象を工夫し、形状やテキストチャ、位置情報などのコンテキストを考慮できる手法を提案した [11]。Winn 等はオクルージョンなど物体間のレイアウトに着目した手法を提案した [15]。Tu 等は CRF を用いずにコンテキストを捉える手法 Auto-Context を提案した [13]。

特徴量の検討では様々な特徴量が検討されている。Scroff 等は、カラーヒストグラム、Texton、HOG 特徴量などの物体認識において一般的に用いられる特徴量の

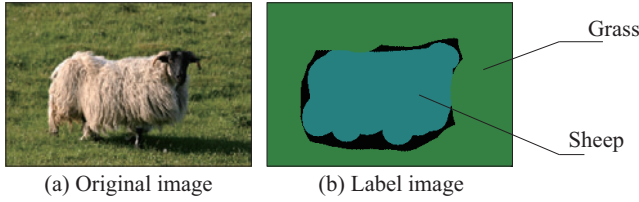


図 1 シングルラベル問題 (MSRC21 データセット [11])

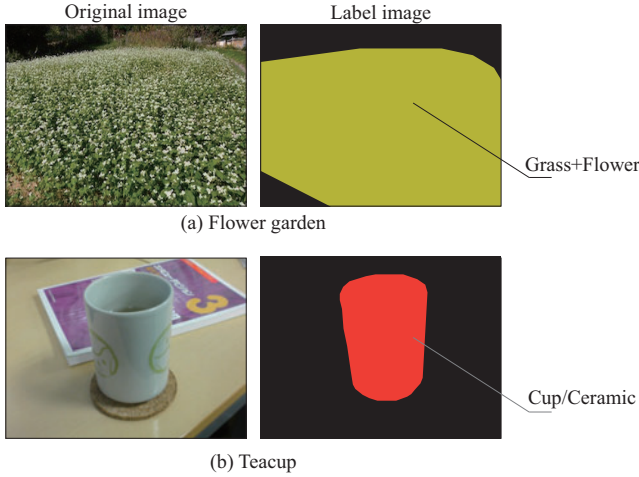


図 2 マルチラベル問題

中から最適な特徴量の検討を行った [9]. Shotton 等は, Texton 特徴量の算出コストを大幅に軽減するために, Random Forests [1] を用いて画像パッチ内の画素を正規化を行わずにテクスチャ解析できる特徴量を学習する Semantic Texton Forests を提案した [10]. また Peter 等は Semantic Texton Forests を改良しラベルの構造を考慮した方法により学習する手法を提案した [5].

上記の手法は, 大量の画素レベルの教師信号付きデータが必要であり, その作成にかかるコストは非常に膨大となる, そこで, 近年では, 学習方法の工夫によりピクセルレベルのラベルが必要ないの簡易化が検討されている. Tighe 等はキーワード付き画像を大量に用意しておき, リトリバル問題とセグメンテーション問題を組み合わせた手法を提案した [12]. また, Vezhnevets 等は Geometric Context と Multiple Instance Learning, Multi-Task Learning を取り入れることで, 画像に対してのみラベルが付けられている弱教師あり学習に展開した [14].

2.2 問題設定

学習を用いるセマンティックセグメンテーション手法では, 図 1 に示すような元画像と教師信号となるラベル画像が対となった学習データを必要とする. ラベル画像とは, 図 1(b) に示すように, 画素毎に属しているカテゴリのラベルが割り当てられたものである. ここで, あるシーンで撮影された 2 枚の画像を図 2(a)(b) に示す. 図 2(a) は花畑が撮影された画像である. この画像に対して,

「花卉」というカテゴリと「葉」というカテゴリを対象とした場合, 花畑はそれらの組み合わせで表現することが可能である. 次に, 図 2(b) は湯のみが撮影された画像である. これに対し, 「コップ」というカテゴリと「陶器」というカテゴリを対象とした場合, それらを組み合わせで陶器製のコップとして湯のみを表現することが可能である. これら 2 つの例は, 同一領域に複数のラベルが割り当てられるマルチラベルの例, しかし前者が構成されるパーツを示しているのに対し, 後者は素材を示している点で異なる.

本稿では, 図 2(a) のようなパーツを表すマルチラベル問題を対象とし, セグメンテーションを行うことを目的とする.

3. Semantic Texton Forests によるセマンティックセグメンテーション

ここでは, Semantic Texton Forests を用いたセマンティックセグメンテーションについて述べる. 図 4 に流れを示す. Semantic Texton Forests は, セマンティックセグメンテーション, 画像識別のための有効な特徴抽出法として提案された [10]. まずはじめに Semantic Texton Forests により画像を識別する. Semantic Texton Forests は学習画像により予め作成しておく. 識別結果をその画像の特徴として, もう一度識別器により各画素を識別することでセマンティックセグメンテーションを行う.

3.1 Semantic Texton Forests

従来法である単一ラベルを用いる Semantic Texton Forests は, Random Forests アルゴリズム [1] により, 大きさ $d \times d$ ピクセルの画像パッチ $\mathbf{p}$ を学習することにより作成する. 画像パッチ $\mathbf{p}$ に対し教師信号 $l_{\mathbf{p}}$ が付けられている. 特徴量の算出は画素参照により行う. 画像パッチ $\mathbf{p}$ から以下の 4 つの特徴量を算出する. $p_{x,y,b}, p_{x_1,y_1,b_1} + p_{x_2,y_2,b_2}, p_{x_1,y_1,b_1} - p_{x_2,y_2,b_2}, |p_{x_1,y_1,b_1} - p_{x_2,y_2,b_2}|$ p は画素を意味している. また x, y は画像パッチ $\mathbf{p}$ の中心からの相対距離, b はカラーチャンネルである.

Random Forests アルゴリズムは, 複数の決定木を作成する. 未知入力 $\mathbf{p}$ に対する, 複数の決定木 T から算出したクラス分布 $P_t(c|\mathbf{p})$ の平均を次式により求め, 識別を行う.

$$P(c|\mathbf{p}) = \frac{1}{T} \sum_{t=1}^T P_t(c|\mathbf{p}) \quad (1)$$

複数の決定木を構築するために学習データをランダム選択することにより作成した木の数 T と同じ数のサブセットを作成する. それぞれのサブセットに対しひとつの決定木を作成する. 各決定木の作成は, 分岐条件を持った分岐ノードの作成を情報利得を基準に繰り返すことにより行う. ここで, 分岐条件は特徴量 f に対し, しきい

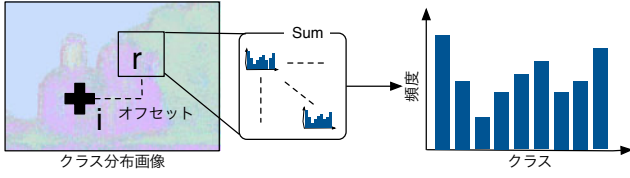


図3 最終識別器による特徴抽出

値 t を用い, $f < t$ ならば左のノードに分岐, $f > t$ ならば右のノードに分岐することとする. この f と t の対を $z = (f, t)$ とし, 現在のノードを n , 学習サンプルを $Q_n = \{\mathbf{p}\}$ として, 左と右にわかれるサンプルのサブセットを以下により得る.

$$Q_l(z) = \{\mathbf{p} \in Q_n | f(\mathbf{p}) < t\} \quad (2)$$

$$Q_r(z) = Q \setminus Q_l(z) \quad (3)$$

そして, 複数の候補をランダムで用意し, その中から最も情報利得 G が大きくなる z を算出し, z を分岐ノードの分岐条件とする.

$$z^* = \arg \max_z G(z) \quad (4)$$

$$G(z) = -\frac{|Q_l|}{|Q_n|} E(Q_l) - \frac{|Q_r|}{|Q_n|} E(Q_r) \quad (5)$$

$G(z^*)$ が十分大きくなるか, 深さが最大値に達したら, 末端ノードを作成する. ここで, E は情報エントロピーである. 学習サンプル Q_n のパッチ $\mathbf{p}$ のラベル l_p を用いてクラスの分布 $P(c|n)$ を作成する.

3.2 クラス分布画像と最終識別器

ここでは, クラス分布画像, 最終識別器の学習法について述べる.

クラス分布画像 まず, クラス分布画像の算出法について述べる. 画素 i を中心に画像パッチ $\mathbf{p}$ を切り出し, Semantic Texton Forests により識別結果 $P_i(c|\mathbf{p})$ を得る. これを画像中の全画素に対し行うことですべての画素に対してクラス分布を得ることができ, それを保持することによりクラス分布画像を作成する.

最終識別器 最終識別器では, カテゴリの位置関係を捉えることにより, 画素を再度識別する. カテゴリの位置関係を捉えるために, クラス分布画像を用いて特徴表現を行う. 図3のように, ある画素 i に対しオフセットした位置の矩形領域 r のクラス分布の合計値 $P(c|r) = \sum_{i \in r} P_i(c|\mathbf{p})$ を特徴量として利用する. オフセットの方向や距離, 矩形の大きさなどは最大値を任意に定め学習により選択する. 最終識別器は, 学習画像により予め作成しておく. 学習アルゴリズムには Random Forests を用いる.

4. 提案手法

Semantic Texton Forests を用いたセマンティックセグメンテーション手法は, 高速かつ高精度な動作を実現で

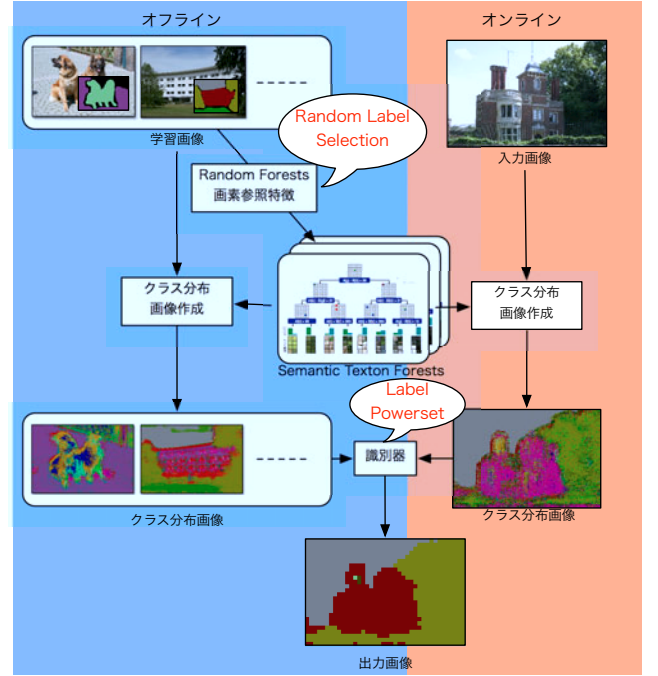


図4 Semantic Texton Forests によるセマンティックセグメンテーションの流れと提案手法による改良点

きるが, マルチラベルには対応していない. そこで, 従来法を改良し, マルチラベルに対応したセマンティックセグメンテーションを実現する. 図4に, 提案手法による改良点を示している. 改良を行うのは, Semantic Texton Forests, 最終識別器に用いる学習法である. Semantic Texton Forests の学習には, 本稿にて提案する Random Label Selection を導入した Random Forests を用い, 最終識別器の学習には, 既存の学習手法をマルチラベルに展開できる Label Powerset 法 [8] を用いる. 以下でそれらの詳細について述べる.

4.1 Random Label Selection を導入した Random Forests

Random Forests は多種多様な問題に応用され, 派生手法も提案されている. しかし, マルチラベルに対応した例はない. そこで, 本稿では, Random Label Selection というアイデアを取り入れ, マルチラベルの学習サンプルを学習できるようにする. ここで, 表記法について述べる. サンプル $\mathbf{p}$ に対し, c をクラスとして, 教師信号の集合 $L_{\mathbf{p}} = \{c | (L_{\mathbf{p}}) > 0\}$, Random Label Selection により選択した教師信号を格納するための $L_{\mathbf{p}}^{test}$ が与えられる. 全学習サンプルを $Q = \{\mathbf{p}\}$, 各決定木を作成するためのサブセットを $Q_s \in Q$. 各ノードで学習に用いるサンプルを $Q_n \in Q_s$ とする. つまり, これらは次の関係にある $Q_n \in Q_s \in Q$.

図5に Random Label Selection を導入した Random Forests のアルゴリズムの模式図を示す. 分岐ノード, 末端ノードそれぞれにおいて新たな処理を加える. 以下に

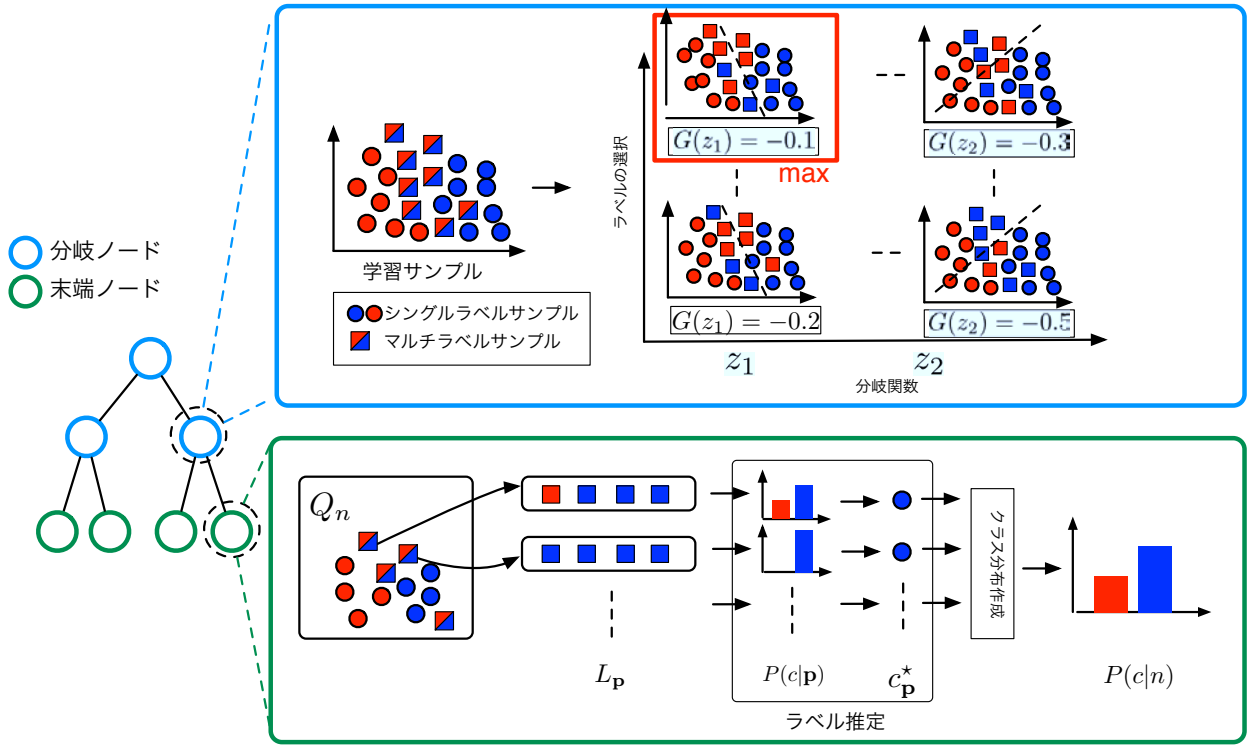


図 5 Random Label Selection を用いた Random Forests

て各処理のアルゴリズムを述べる。

分岐ノード アルゴリズム 1 に分岐ノードにおける処理を示す。本アルゴリズムにおいて、分岐ノードでは分岐関数 z だけでなく、マルチラベルサンプルのラベルも選定する。まず、通常の Random Forests と同様に分岐関数を選択する。次に、本手法では新たにサンプル $\mathbf{p} \in Q_n$ のラベル $L_{\mathbf{p}}$ から、選定に用いるラベル $c_{\mathbf{p}}^{test} \in L$ を以下の式により得る。

$$c_{\mathbf{p}}^{test} = \begin{cases} \text{rand}(L_{\mathbf{p}}) & |(L_{\mathbf{p}})| > 2 \\ \text{single}(L_{\mathbf{p}}) & \text{otherwise} \end{cases} \quad (6)$$

関数 rand は $L_{\mathbf{p}}$ のラベルをランダムに選択し、出力する。関数 single はシングルラベルサンプルのラベルを出力する。情報利得の算出には、このラベル $c_{\mathbf{p}}^{test}$ を用いる。 $c_{\mathbf{p}}^{test}$ の選定は複数回行い、その回数を β とし、本アルゴリズムのパラメータとする。これらの処理により、最も情報利得が高くなった $c_{\mathbf{p}}^{test}$ を $L_{\mathbf{p}}^{test}$ に追加し、特徴量としきい値を用い、分岐ノードを作成する。これを繰り返すことにより、各分岐ノードにつきひとつのラベルが $L_{\mathbf{p}}^{test}$ に追加されていくことになる。

末端ノード アルゴリズム 2 に末端ノードにおける処理の流れを示す。末端ノードでは、分岐されてきたサンプル Q_n を用いてクラス分布を算出する。ここで、マルチラベルサンプルに対しては、複数あるラベルから最もそのサンプルに相応しいラベルを選択する。そのために、 $L_{\mathbf{p}}^{test}$ に基づき、マルチラベルサンプルのラベルの推定を行う。推定は次式により行う。

Algorithm 1: 分岐ノードの処理

Require: 学習サンプル Q_n

Require: 分岐関数の選定回数 α

Require: ラベルの選択回数: β

For $i = 1, \dots, \alpha$

・分岐関数をランダムに選択

For $j = 1, \dots, \beta$

For Each $\mathbf{p} \in Q_n$

$$c_{\mathbf{p}}^{test} = \begin{cases} \text{rand}(L_{\mathbf{p}}) & |(L_{\mathbf{p}})| > 2 \\ \text{single}(L_{\mathbf{p}}) & \text{otherwise} \end{cases}$$

End For

・情報利得算出: 分岐関数 $c_{\mathbf{p}}^{test}$ を保持

End For

For Each $\mathbf{p} \in Q_n$

・選定ラベルを追加 $L_{\mathbf{p}}^{test} \leftarrow c_{\mathbf{p}}^{test}$

End For

$$c_{\mathbf{p}}^* = \begin{cases} \arg \max_c (P(c|\mathbf{p})) & |(L_{\mathbf{p}})| > 2 \\ \text{single}(L_{\mathbf{p}}) & \text{otherwise} \end{cases} \quad (7)$$

$P(c|\mathbf{p})$ は、 $\mathbf{p}$ に対するクラスらしさであり、 $L_{\mathbf{p}}^{test}$ により算出した、クラスの選択頻度を用いて算出する。これにより算出した $c_{\mathbf{p}}^*$ を用いて末端ノードのクラス分布 $P(c|n)$ を作成する。

以上が、Random Label Selection を用いた Random Forests の学習アルゴリズムである。分岐ノードにおいては、ラベルをランダムに選択すること最も分岐に貢献したラベルを保存した。末端ノードにおいては、分岐ノードにて保存されてきたラベルの情報を用い、最もらしい

Algorithm 2: 末端ノードの処理

Require: 学習サンプル Q_n **For Each** $\mathbf{p} \in Q_n$

$$c_{\mathbf{p}}^* = \begin{cases} \arg \max_c (P(c|\mathbf{p}) \mid |(L_{\mathbf{p}})| > 2) \\ \text{single}(L_{\mathbf{p}}) & \text{otherwise} \end{cases}$$

End For $c_{\mathbf{p}}^*$ を用いてクラス分布 $p(c|n)$ 作成

ラベルを推定させクラス分布を作成した。これらにより、マルチラベルへの対応が可能となった。

4.2 Label Powerset を用いた最終識別器のマルチラベルへの対応

未知サンプルに対し、マルチラベルを結果として出力するためには、識別器の取得に対し、しきい値処理を行い、しきい値以上のクラスを出力することや、事後確率から条件付き情報エントロピーを算出することによりそれを基準にすることが考えられる。しかしながら、しきい値などの定数を利用しなければならないため、動的な環境への対応は難しい。そこで、本稿ではそのような定数は利用せずにマルチラベルに対応できる手法として、Label Powerset [8] を用いる。Label Powerset は表 1 のようにマルチラベルに対して擬似クラスを生成することにより、マルチラベルをひとつのクラスとして考える。これを前処理として行い、本手法においては Random Forests により学習することで、擬似クラスによりマルチラベルに対応することが可能となる。

表 1 Label Powerset の利用例

Sample	Label	Label after processing
1	$\{c_1, c_4\}$	$\{c_{1,4}\}$
2	$\{c_3, c_4\}$	$\{c_{3,4}\}$
3	$\{c_1\}$	$\{c_1\}$
4	$\{c_2, c_3, c_4\}$	$\{c_{2,3,4}\}$

4.3 マルチラベルに対応したセマンティックセグメンテーション

ここまでで、各識別器のマルチラベルへの拡張法について述べた。ここでは、図 4 の流れに従い、識別時におけるマルチラベルに対応したセマンティックセグメンテーションの流れについて述べる。未知画像が入力されると、Random Label Selection を導入した Random Forests アルゴリズムにより構築した Semantic Texton Forests により識別を行い、クラス分布画像を得る。次に、Label Powerset を導入した最終識別器により、識別を行うことにより結果画像を得る。以上がマルチラベルに対応したセマンティックセグメンテーションの流れである。

ここで、提案手法における最終識別器の特徴抽出法について 3. 節にて述べた文献 [10] で用いられる方法とは

異なる方法を利用するため詳細を述べる。本稿においては、図 3 のようにオフセットした位置からの特徴抽出は行わず、注目画素の周囲の情報のみを利用する。また、矩形サイズも学習による選択は行わず固定値を用いる。

5. 実験および考察

ここでは、提案手法の有効性を示すための評価実験と考察について述べる。まず、実験概要について述べ、次の評価結果を述べる。

5.1 実験概要

Random Label Selection を導入した Random Forests にて学習を行う Semantic Texton Forests, 最終識別器の学習パラメータを表 3 に示す。最終識別器に用いる Random Forests には Random Label Selection を用いないため、ラベル選択回数 β はなしとしている。

表 3 提案手法の学習パラメータ

項目	Semantic Texton Forests	最終識別器
決定木の数	5	5
決定木の深さ	10	10
分岐閾数候補数	2000	300
ラベル選択回数 β	2	なし
サブセットの割合	20%	20%

実験には、Flickr よりダウンロードした VGA サイズの画像をラベリングしたデータを用いる。図 6 にデータセットの一部を示す。対象カテゴリは、Face, Body, Tree, Grass, Sky, Flower, Neon の 7 カテゴリ、Grass + Flower, Sky + Tree のマルチラベル 2 カテゴリである。実験データは、すべて合わせて 1168 枚の画像があり、学習用に 587 枚の画像、評価用に 581 枚の画像を利用する。

本実験において、評価する項目は以下の 3 つである。

- (1) 各カテゴリの識別率
- (2) Random Label Selection の有効性の評価
- (3) パラメータ β の識別結果への影響

まず、1 については、識別結果の各画素とラベル画像を比較し、Confusion Matrix を用いて結果を示す。これにより、カテゴリ間の関係を交えた評価を行う。次に 2、3 については、各クラス毎に再現率 (Recall) を算出し、平均した値を利用する。正解画像のクラス c の領域をセグメンテーション結果においてクラス c と出力できた面積を True Positive (TP) とし、正解画像のクラス c の領域のうち他のクラスと間違われた領域を False Negative (FN) とする。この時、クラス c に対する $Recall_c$ は次のように求める $Recall_c = \frac{TP}{TP+FN}$ 。そして、クラスの数 N とし、平均再現率 (Average Recall) を以下の式により算出する。

$$\text{Average Recall} = \frac{\sum_c Recall_c}{N} \quad (8)$$



図 6 データセット

表 2 Confusion Matrix

		Predicted								
		sky	grass	neon	tree	face	body	flower	grass+flower	sky + tree
Ground Truth	sky	0.84	0.00	0.00	0.00	0.03	0.00	0.11	0.00	0.01
	grass	0.03	0.93	0.00	0.02	0.00	0.00	0.01	0.01	0.00
	neon	0.00	0.13	0.56	0.05	0.02	0.07	0.04	0.04	0.09
	tree	0.00	0.12	0.03	0.59	0.01	0.03	0.00	0.11	0.10
	face	0.00	0.03	0.01	0.10	0.84	0.01	0.01	0.00	0.00
	body	0.02	0.00	0.02	0.02	0.18	0.57	0.07	0.11	0.00
	flower	0.04	0.00	0.03	0.00	0.14	0.04	0.66	0.07	0.01
	grass+flower	0.09	0.02	0.11	0.03	0.02	0.03	0.15	0.64	0.00
	sky+tree	0.02	0.01	0.04	0.21	0.00	0.04	0.02	0.02	0.65

5.2 各カテゴリの識別率の評価とマルチラベルの有効性の確認

提案手法による Confusion Matrix を表 2 に示す。結果より、マルチラベルのサンプル Grass+Flower, Sky + Tree はそれぞれ、サブカテゴリ Grass, Flower, Sky, Tree に混同されやすいことが分かる。しかし、マルチラベルカテゴリのサブカテゴリとして利用されている、シングルラベルカテゴリ Grass, Flower, Sky, Tree に主観を置くと、マルチラベルのカテゴリにそれほど混同されないことが分かる。サンプルの数が異なることも影響することも考えられるが、提案手法はシングルラベルカテゴリの識別結果にあまり影響を与えずに、マルチラベルカテゴリの識別が可能であることが分かる。

識別結果の例を図 9 に示す。また、マルチラベルに対応したことによる効果を確認するために、シングルラベルのみを考慮した結果を STF_s として示す。(a) のシングルラベルのサンプルでは従来の STF_s を用いた結果と提案手法を用いた結果の差が無いことが分かり、提案手法は、従来法と同等のシングルラベルのセグメンテーションを行えることが分かる。また、(b) のマルチラベルのサンプルの grass + flower の例では、STF_s は花畑を flower と grass に分割しているのに対し、提案手法では

grass+flower として花畑を分割することができていることが分かる。さらに、tree+sky を扱った例では、STF_s は木の枝の部分を tree でも sky でも無く、flower として誤認識しているのに対し、提案手法では、マルチラベルを考慮することで、sky+tree として捉えることができています。これらの結果により、マルチラベルに対応したことによる有効性を確認できたといえる。

5.3 Random Label Selection の有効性の評価

Random Label Selection を導入した提案手法の有効性を確認するために、比較実験を行う。比較対象は、マルチラベルサンプルを学習対象として利用せずに、通常の Random Forests により学習した Semantic Texton Forests を利用し、Label Powerset を用いた最終識別器によりセグメンテーションを行う手法である。この手法も、最終識別器に Label Powerset を用いていることにより、マルチラベルの結果を得ることが可能である。本実験ではこの手法を従来法とする。評価値には平均再現率を用いて比較を行うが、効果が分かりやすいように、シングルラベルカテゴリとマルチラベルカテゴリに分けて平均再現率を算出することで比較を行う。

図 7 に結果を示す。結果より、Random Label Selec-

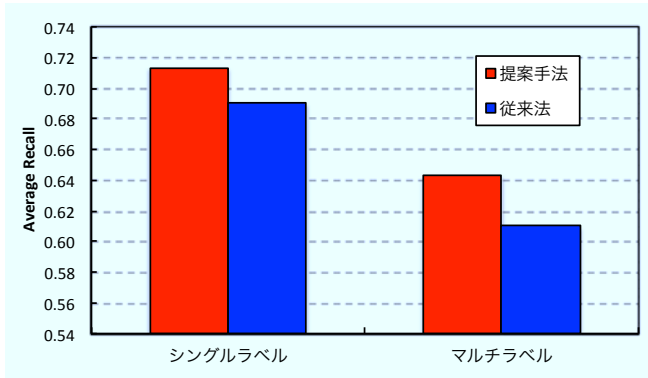


図7 提案手法と従来法の比較

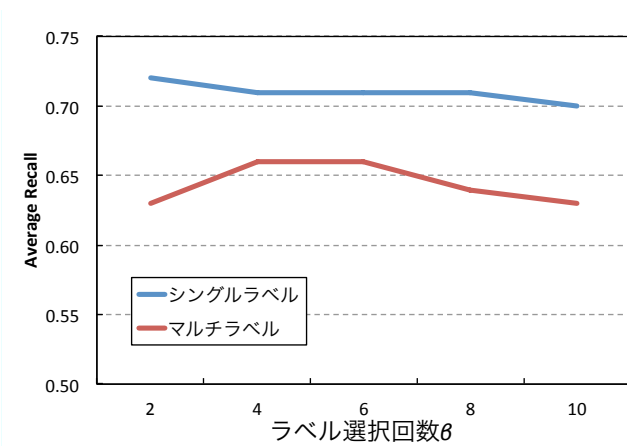


図8 パラメータ β に対する平均再現率

tionを導入することで、シングルラベルカテゴリにおいて2%、マルチラベルカテゴリにおいては、4%の平均再現率の向上が見られた。これにより、Random Label Selectionを用いてマルチラベルサンプルを学習することにより、マルチラベルカテゴリだけでなく、シングルラベルカテゴリも捉えやすい特徴表現が可能になっていると考えられる。これには、Random Label Selectionを用いたRandom Forestsは分岐ノード、末端ノードの処理を通し、マルチラベルサンプルを用いてシングルラベルカテゴリを学習している点が関係していると考えられる。マルチラベルサンプルがシングルラベルカテゴリのみでは表現できない特徴を持ち、識別境界を変化させる働きがあったためこのような効果が得られたと考えられる。

5.4 ラベル選択回数 β の傾向調査

提案手法の学習パラメータであるラベル選択回数 β の傾向を調査する。 β を2から10まで2刻みで変化させ、マルチラベルカテゴリとシングルラベルカテゴリの平均再現率を求める。図8に結果を示す。結果より、 β を変化させても再現率にはほぼ変化が無かった。この理由として、今回の問題はマルチラベルカテゴリのラベルの組み合わせ数は、最大2カテゴリ (Tree+Sky = 2, flower + Tree = 2カテゴリ) であることがあげられる。そのた

め、それほど多くのラベルの選定が必要なかったと考えられる。また、他の要因として、アルゴリズム1に示したように、ラベルの選定は分岐関数の選択に対し入れ子になっている点が挙げられる。分岐関数選定の際に類似した分岐関数が選定に利用された場合、利用した β よりも多くの選定を行ったのと類似した効果が得られる。

以上により、 β はラベルの組み合わせ数、分岐関数の選定回数を考慮して決定する必要があると考えられる。図8から今回の問題では実験に用いた最も低いパラメータ2でも十分であると言える。

6. おわりに

本稿では、Random Label Selection というアイデアを元に、マルチラベルに対応したセマンティックセグメンテーションを実現した。提案手法はSemantic Texton ForestsをRandom Label Selectionを導入したRandom Forestsにより学習し、後段の識別器にはLabel Powerset法を用いた。独自に構築したデータセットを用いた評価実験により、平均再現率約0.7にてセグメンテーションが可能であることを示した。今後は、マルチラベルカテゴリの組み合わせ数、カテゴリ数を増やした場合についても検討を行う。

文 献

- [1] L. Breiman. Random forests. *Machine learning*, 45(1):5–32, 2001.
- [2] B. Fulkerson, A. Vedaldi, and S. Soatto. Class segmentation and object localization with superpixel neighborhoods. In *International Conference on Computer Vision*, pages 670–677, 2009.
- [3] J. M. Gonfaus, X. Boix, J. van de Weijer, A. D. Bagdanov, J. Serrat, and J. González. Harmony potentials for joint classification and segmentation. In *Computer Vision and Pattern Recognition*, pages 1–8, San Francisco, California, USA, 2010.
- [4] X. He, R. S. Zemel, and M. A. Carreira-Perpinan. Multiscale conditional random fields for image labeling. In *Computer Vision and Pattern Recognition*, 2004.
- [5] P. Kotschieder, S. R. Bulò, H. Bischof, and M. Pelillo. Structured class-labels in random forests for semantic image labelling. In *International Conference on Computer Vision*, pages 2190–2197, 2011.
- [6] D. Larlus and F. Jurie. Combining appearance models and markov random fields for category label object segmentation. In *Computer Vision and Pattern Recognition*, 2008.
- [7] S. Nowozin, P. V. Gehler, and C. H. Lampert. On parameter learning in crf-based approaches to object class image segmentation. In *European conference on Computer vision*, 2010.
- [8] J. Read. A pruned problem transformation method for multi-label classification. In *Proc. New Zealand Computer Science Research Student Conference*, pages 143–150, 2008.
- [9] F. Schroff, A. Criminisi, and A. Zisserman. Object class segmentation using random forests. In *British Machine Vision Conference*, 2008.
- [10] J. Shotton, M. Johnson, and R. Cipolla. Semantic texton forests for image categorization and segmentation. In *Computer Vision and Pattern Recognition*,

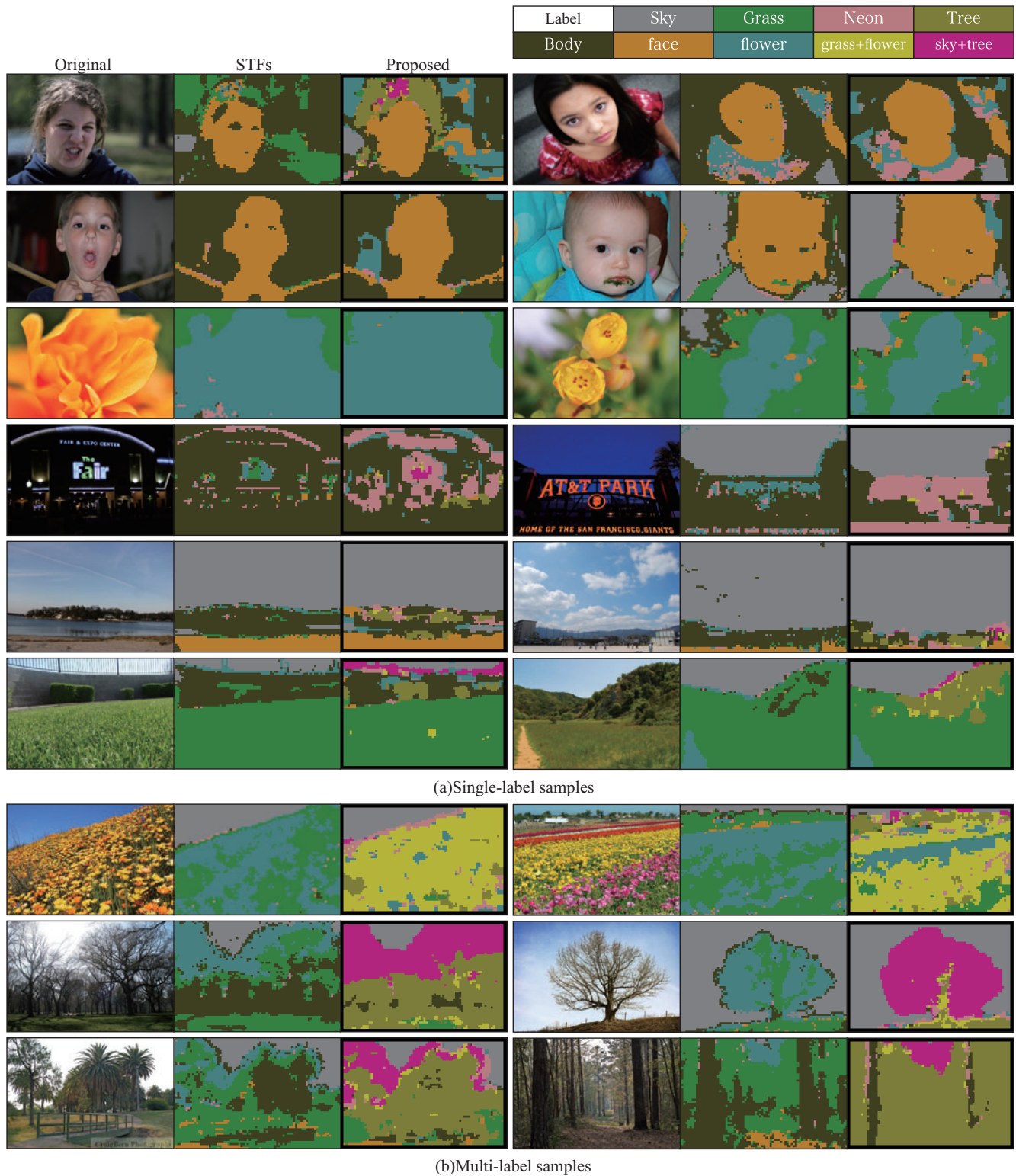


図 9 セグメンテーション結果

- 2008.
- [11] J. Shotton, J. Winn, C. Rother, and A. Criminisi. Textonboost: Joint appearance, shape and context modeling for multi-class object recognition and segmentation. In *European Conference on Computer Vision*, 2006.
 - [12] J. Tighe and S. Lazebnik. Superparsing: Scalable nonparametric image parsing with superpixels. In *European Conference on Computer Vision*, 2010.
 - [13] Z. Tu. Auto-context and its application to high-level vision tasks. In *Computer Vision and Pattern Recognition*, 2008.
 - [14] A. Vezhnevets and J. M. Buhmann. Towards weakly supervised semantic segmentation by means of multiple instance and multitask learning. In *Computer Vision and Pattern Recognition*, pages 3249–3256, 2010.
 - [15] J. Winn and J. Shotton. The layout consistent random field for recognizing and segmenting partially occluded objects. In *Computer Vision and Pattern Recognition*, 2006.