社団法人　電子情報通信学会
THE INSTITUTE OF ELECTRONICS,
INFORMATION AND COMMUNICATION ENGINEERS

信学技報
TECHNICAL REPORT OF IEICE.

ラーメン vs カレー: 2年分のログデータと高速食事画像認識エンジンを用いたTwitter食事画像分析とデータセット自動構築

河野　憲之†　　柳井　啓司†

† 電気通信大学大学院　〒 182–8285 東京都調布市調布ヶ丘 1–5–1
E-mail: †kawano-y@mm.cs.uec.ac.jp, ††yanai@cs.uec.ac.jp

あらまし　多くの人々が Twitter を利用するようになり，大量に投稿されたツイートを通して人々の行動や考えを分析することが可能となった．ツイートには画像が付与されたものも多く，特に昼食時，夕食時には，食事の画像が大量にツイートされる．そこで，本稿では 2011 年 5 月から 2013 年 8 月の 2 年 4ヶ月の間に収集した約 10 億件の画像付きツイートに対して，食事キーワード検索と高速食事画像認識エンジンを用いて，100 種類の食事画像を抽出する実験を行った結果を報告する．実験では，食事画像ランキング，一部の食事カテゴリについてサンプリングによる抽出精度評価，また位置情報食事画像ツイートを用いた「ラーメン」と「カレー」に関する地域分布の分析を行った．またさらに，我々が構築した 100 種類の食事画像データセットを自動的に拡張するためのフレームワークについても述べる．100 種類食事画像データを利用して構築した食事画像判定エンジンと，Amazon Mechanical Turk を利用したクラウドソーシングを用いて，キーワードを与えるのみで，自動的に新しい食事カテゴリのバウンディングＢＯＸ付きの画像データセットを構築する．実験では，手動で作成した既存の食事画像データセットのサブセットとの認識精度の比較を行う．

キーワード　Twitter，食事画像認識，食事画像判別エンジン，データセット自動構築，クラウドソーシング

1. ま え が き

近年，スマートフォンやタブレット端末の普及により，Twitter(注1) へ場所に依らず気軽にコメントを発信できるようになった．投稿される内容は様々であるが，特に昼食時，夕食時には，外食や自炊した食事を撮影し，ツイートするケースが多く見受けられる．ツイートのテキスト，また画像を分析し事象や流行を読むことは行われ，成果が得られている [1] [2]．しかしながら，量の多さから，ツイートの画像を分析することはクラスタリング程度の簡単な処理に留まっており [2]，近年盛んに研究されている一般物体認識手法を用いてツイートに含まれる画像を分析することは行われていない．

我々は以前，スマートフォン上でリアルタイムに食事画像を認識し記録をとるシステムを提案した [3] [4]．また，柳井は Twitter のストリームを監視し，Twitter に投稿された画像をリアルタイムにマッピングし，蓄積するシステムを提案した [5]．そこで，本稿では，スマートフォン向けのリアルタイムの食事画像認識エンジンを PC に実装し，Twitter ストリーム監視システムで蓄積した画像データを分析することを行う．PC に実装した 100 種類食事画像認識エンジンは上位 5 位以内の候補に約 80%の精度で食事画像を分類可能で，1 枚の認識の実行時間は 0.024 秒である．これを利用して，2011 年 5 月から 2013 年 8 月の 2 年 4ヶ月の間に収集した約 10 億件の画像付きツイートから食事キーワード検索で選ばれた約 170 万枚の食事候補画像について，100 種類の食事画像を抽出する実験を行う．実験では，食事画像ランキング，一部の食事カテゴリについてサンプリングによる抽出精度評価，また位置情報食事画像ツイートを用いた「ラーメン」と「カレー」に関する地域分布の分析を行った．

また，我々は以前より食事画像認識の研究を行っており，食事画像認識のためのデータセットを自力で収集し，GroundTruth としてバウンディングボックスを付与してきた．また，日本でよく食べる食事を収集してきた．だが，食事の数はもっと多く存在しているため，現在認識対象にしている 100 種類では，認識対象が不足している．そこで，本稿では，Web から食事画像を収集し，画像に映っている食品にバウンディングボックスを付与する作業を Amazon Mechanical Turk を利用したクラウドソーシングを用いることで自動的に行い，食事画像データセットの自動構築を試みる．加えて，日本食以外の海外でよく食べられると考えられる食事についても実験を行った．

(注1)：https://twitter.com

2. 関 連 研 究

2.1 Twitter に投稿されたツイートの分析

Twitter は，人々がリアルタイムにツイートをする性質をもつ．この性質を利用し，榊ら [1] は，ユーザをソーシャルセンサと捉え，モデル化することによりリアルタイムに地震などの自然災害を検出した．金子ら [2] は，一定の領域ごとにイベントの存在を求め，クラスタリングにより，視覚的なイベントを検出した．柳井 [5] は，Twitter ストリームを監視しリアルタイムに位置情報付き画像を取得し，マッピング，またそれらのクラスタリングを行った．

本稿では，ツイートの分析対象を食事画像とし，過去 2 年間の Twitter のログを分析し,「ラーメン」と「カレー」に関する地域分布の分析を行った．

2.2 クラウドソーシングとコンピュータビジョン

クラウドソーシングを用いて，データセットの作成はよく行われるようになった．クラウドソーシングによりアノテーションされたデータを用い，物体検出器の学習を行った研究を紹介する．Vijayanasimhan [6] らは，クラウドソーシングを用い，アクティブラーニングすることで物体検出の精度を向上させた．高速化手法として，hyperplane-hashing により，物体検出に用いる SVM の超平面に近いサンプルをクラウドソーシングに用いている．大澤ら [7] も，同様にループ学習により食事検出器の自動構築を試みている．

本稿では，収集する対象を食事に限定してクラウドソーシングを用い，バウンディングボックスの付与を行い，データセットの自動構築を目標とする．

2.3 クラウドソーシングとその結果の分析

ワーカーは必ずしも正確に依頼したタスクを実行してくれないため，クラウドソーシングで得られた結果は信頼できない場合がある．そこで，Raykar [8] らは，二値分類，N 値分類，採点のラベル付けタスクにおいて，ワーカーのランキング手法を提案した．評価値は [0,1] 間になるように設計されているため，閾値設定が容易になっている．

本稿では，主にバウンディングボックスを付与するというタスクであり，Raklar らとのタスクとは性質が異なっている．画像中にバウンディングボックスを描画することは，自由度が高いことに注目することで，スパマーの結果の除去を行う．

3. 食事画像認識エンジン

3.1 100 種類食事画像認識エンジン

我々が以前提案したスマートフォン上で食事画像認識を高速に行う認識エンジンについて説明する．認識する領域は大きい場合は 30K にリサイズした．局所特徴量は，HOG Patch [9] と Color Patch を用いた．HOG Patch は 32 次元 (2x2，8 方向)，Color Patch は 24 次元 (2x2，RGB の平均と分散) とした．各局所特徴量に PCA を適用し，それぞれ 24 次元にした．GMM のコンポーネント数を 16 として，GMM を作成した．そして，Fisher Vector (FV) [10] [11] により，画像特徴量とした．SVM により学習を行い，late fusion で結合した．認識時の演算回数を少なくする，クアッドコアで並列処理することで高速な認識を行った．

本稿では，スマートフォンでなく，一般的なクアッドコアの linux マシンに認識エンジンを組み込んだ．スマートフォンと比較し，メモリが十分にあり，処理を認識のみにあてることができるため，計算コストを高め，認識精度を高めた．認識する領域は中央 64%とし，大きい場合は 50K にリサイズした．同様に 2 スケール 6 ピクセルごとの dense sampling により局所特徴量を抽出し，PCA を適用したが，次元数はそれぞれ保った．また，GMM のコンポーネント数を 64 に変更した．そして，SPM [12] を適用した．1x1，2x2 に分割したサブ領域から 5 つの FV を作成した．最終的に，HOG-FV は 20480 次元，Color-FV は 15360 次元となった．認識精度は，交差検定により評価を行い，100 種類の食事において，正解領域が与えられたとき，59.6%の分類率を示した．認識時間は，コンパイラは Intel C++ Compiler 11.0，CPU は Intel Core i7-3770K 3.50GHz を用い実時間を計測した．640x480 の画像に対し 0.024 秒であった．高速に食事画像認識を行えていることがわかる．

3.2 食事画像判別エンジン

キーワードにより Web から収集しただけの画像には，ノイズが多く含まれる．そのため，収集した画像がキーワードで指定した画像であるかを判定する必要がある．

本稿では，収集した画像のフィルタリングに，識別性能の高さから画像認識でよく用いられる SVM を用いた．まず，既存食事画像データセットをいくつかに分割した．分割方法は，様々存在するが，本稿では，1-vs-rest 線形 SVM の混合行列を用いた．次のステップで SVM を用いるため自然な分割方法であると考える．画像特徴量には，HOG-FV，Color-FV を使用した．交差検定により評価し，混合行列を得た．他のカテゴリと混合が多い条件として，正しく分類できた数の x% (x=40) 以上あるカテゴリに誤分類されていた場合，それらのカテゴリ同士を結合する．結合されるカテゴリが存在しない場合は，x を半減させる．これらを再帰的に行い，カテゴリ数を減らし，13 の食事グループが作成された．結合されなかったカテゴリは除去した．そして，ポジティブ画像はそれぞれグループに属する食事画像，ネガティブ画像は後述と同様の方法で，ネガティブとして収集されやすいと想定した語 (「street stall」,「kitchen」,「dinner party」,「restaurant」) で収集し，食事画像は手動で除外し使用した．最終的な評価値は，構築した 13 の識別器の最大反応（最大評価値）を示した値を付与する．この食事画像判別エンジンは，SVM で識別が難しい食事グループを結合し，その食事グループをみわけるため，学習時のノイズデータによる識別性能低下が少なく，他の未知カテゴリの食事画像においても，適用することが可能になっている．Twitter 食事画像分析では，この評価値を用いて食事画像であるかの判定を行った．クラウドソーシングによるデータセット自動構築では，この評価値を食事性として扱い，評価値の高い，より食事画像らしいものを使用した．

4. Twitter 食事画像分析

本節では，前節で述べた高速食事画像認識エンジンの大量データへの応用例として，我々が 2011 年 5 月から 2013 年 8 月の 2 年 4ヶ月間に渡って収集した，Twitter に投稿された画像付きの約 10 億件 (正確には 988,884,946 件) のツイートに対して，100 種類食事画像認識エンジンが対象とする 100 種類の食事の画像分析を行った結果について報告する．

まず 100 種類の食事名のキーワード検索によって，画像を 1,730,441 枚に絞り込んだ．100 種類の食事名キーワードはここでは日本語に限定したため，大幅に枚数が少なくなった．しかしながら，それでもなお 100 万枚を超える画像があり，それらに対してすべて一般物体認識を行うのは一般には時間の掛る処理である．それに対して，今回実装した高速食事画像認識エンジンは 1 枚あたりの認識時間が 0.024 秒程度で極めて高速なため，例えば 100 万枚の画像を認識する場合に，50 台の Quad-core マシンが利用できれば，10 分も掛からずに認識を終えることが可能である．

約 10 億件の画像 URL を含むツイートに対して，100 種類の食事名によってキーワード検索を行いツイートの絞り込みを行い約 170 万枚に絞り込んだ後，100 種類の食事名が含まれるツイートの画像に対して，**食事画像判別エンジン**と **100 種類食事画像認識エンジン**を適用して，画像認識による画像の絞り込みを行う．最終的には，約 10 億件のツイートに含まれる画像のうち，約 47 万枚 (正確には 470,335 枚) が 100 種類の食事画像に相当すると判定された．

具体的には，食事名のキーワードがツイートのメッセージ中に含まれているツイートを検索し，最初に食事画像判別エンジンで食事画像であるかどうかの判定を行う．食事画像と判定された場合は，次に 100 種類食事画像認識エンジンで，食事の種

類の認識を行う．この認識結果の上位 5 位以内に，ツイート中に含まれる食事名と同じ食事カテゴリが含まれていれば，その画像はツイート中に含まれる食事名に対応するとみなすこととする．この結果の精度については，後ほど評価を行う．

こうして選ばれた画像に対して，次の 3 点の分析を行った．(1)100 種類の食事のそれぞれの出現枚数，(2) 一部の食事についてキーワード検索のみ，食事判別エンジンを適用した結果，100 種類認識エンジンを適用した結果，2 つの認識エンジンを同時に適用した結果，の 4 通りについてのそれぞれの食事画像集合の適合率，(3) 位置情報付きツイートに含まれる食事画像を利用し，出現頻度が上位 1，2 位であるラーメンとカレーに関して，その地理的な分布の比較，をそれぞれ行った．

表 1　100 種類の出現頻度ランキング．

1	ラーメン	80021	34	たい焼き	3281	67	干物	563
2	カレー	59264	35	カツ丼	3188	68	シュウマイ	561
3	寿司	25898	36	オムレツ	2592	69	フライドポテト	561
4	つけ麺	22158	37	ビビンバ	2368	70	チャーシューメン	555
5	オムライス	17520	38	スパゲッティ	2171	71	サンドウィッチ	551
6	ピザ	16921	39	たたき	2162	72	冷奴	517
7	餃子	16014	40	肉じゃが	2129	73	筑前煮	352
8	お好み焼き	15234	41	納豆	2094	74	豚カツ	331
9	ご飯	14264	42	ミートソース	1994	75	南蛮漬け	323
10	刺身	13927	43	茶碗蒸し	1843	76	鶏の唐揚げ	314
11	ハンバーグ	11583	44	目玉焼き	1635	77	チンジャオロース	312
12	ステーキ	9503	45	クロワッサン	1579	78	ロールパン	288
13	たこ焼き	9004	46	かけうどん	1500	79	ローストチキン	263
14	チャーハン	8383	47	角煮	1443	80	マカロニサラダ	239
15	焼きそば	7905	48	ちらし寿司	1371	81	煮魚	228
16	おでん	7453	49	豚汁	1229	82	きんぴらごぼう	225
17	トースト	6350	50	生姜焼き	1158	83	天ぷらうどん	213
18	カツカレー	6339	51	ポテトサラダ	1150	84	レーズンパン	205
19	天ぷら	5905	52	卵焼き	1146	85	ゴーヤーチャンプル	198
20	おにぎり	5462	53	うな重	1071	86	グリーンサラダ	145
21	グラタン	5223	54	春巻き	1058	87	中華スープ	141
22	コロッケ	4837	55	酢豚	1049	88	けんちん汁	137
23	シチュー	4797	56	エビフライ	1049	89	鮭のムニエル	96
24	海鮮丼	4730	57	野菜炒め	1040	90	さんまの塩焼	84
25	親子丼	4513	58	エビチリ	1003	91	惣菜パン	76
26	天丼	4464	59	ロールキャベツ	965	92	魚のフライ	72
27	牛丼	4285	60	炊き込みご飯	901	93	野菜天ぷら	71
28	麻婆豆腐	4081	61	ピラフ	891	94	天津麺	69
29	焼き鳥	3829	62	ざるそば	880	95	がんもどき	34
30	ハンバーガー	3662	63	コーンスープ	816	96	鮭の塩焼	25
31	冷やし中華	3473	64	ホットドッグ	795	97	ほうれん草炒め	12
32	すき焼き	3408	65	チキンライス	736	98	魚の照り焼き	3
33	味噌汁	3295	66	ウィンナ	577	99	なすの油味噌	2
						100	ピザトースト	0

図 1　オムレツ画像の例．文字や絵が書かれいることが多い．

表 2　処理ステップ毎の選択枚数とその精度．精度評価は 300 枚のランダムランプリング評価による．

品目	キーワードのみ	食事検出	100 種類認識	検出＋認識
ラーメン	275652 (72.0%)	200173 (92.7%)	84189 (95.0%)	80021 (99.7%)
チャーシュー麺	861 (94.3%)	811 (99.0%)	558 (99.7%)	555 (99.7%)
カレー	224685 (75.0%)	163047 (95.0%)	62824 (97.0%)	59264 (99.3%)
カツカレー	10443 (92.7%)	9073 (98.0%)	6544 (98.7%)	6339 (99.3%)
寿司	86509 (69.0%)	43536 (86.0%)	48019 (72.3%)	25898 (92.7%)
つけ麺	33165 (88.7%)	24896 (96.3%)	28846 (93.7%)	22158 (99.0%)
オムライス	34125 (90.0%)	28887 (96.3%)	18370 (98.0%)	17520 (99.0%)

まず表 1 に 470,335 枚の内訳を示す．結果は，ラーメンとカレーが他を大きく引き離して 1 位，2 位であった．以下，寿司，

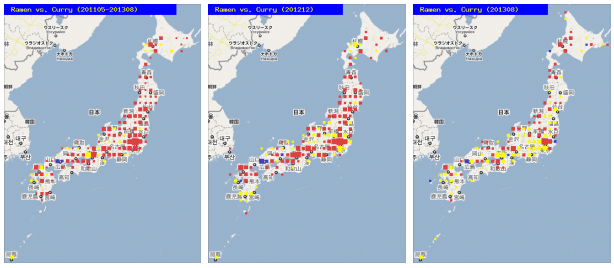

図 2　ラーメン (赤印)，カレー (黄印), お好み焼き (青印) のうち，最もツイートが多かった地域にマークを付けた日本地図．左は 2 年 4ヶ月分，中央は 2012 年 12 月，右は 2013 年 8 月の地域毎の主要食事画像ツイートの分布を示す．冬はラーメンが多く，夏はカレーが多いことが分かる．

つけ麺，オムライスであった．つけ麺はラーメンのバリエーションとも考えられるが，単独のカテゴリとして 4 位になるのは意外であった．また，図 1 に示すようにオムライス画像の約半分は，ケチャップで絵や文字が書かれているオムライスであり，手作りのものが多いのが特徴である．普段良く食べる牛丼やハンバーガーはチェーン店のものが多く，それぞれの食事に個性がないので，20 位以下であまり順位は高くない．一方，ラーメンやカレーは専門店が多く，個性的なものが多いため，ツイートされる画像も多くなっていると考えられる．

表 2 は，上位 5 品目にラーメンの下位カテゴリであるチャーシュー麺，カレーの下位カテゴリであるカツカレーを追加した 7 種類について，キーワード検索のみ，食事判別エンジンを適用した結果，100 種類認識エンジンのみを適用した結果，2 つの認識エンジンを適用した結果，の 4 通りについて，それぞれの食事画像集合の枚数および 300 枚のランダムサンプリング評価による適合率を示す．まず，キーワード検索の結果であるが，下位カテゴリのチャーシュー麺，カツカレーは 94.3%, 92.7%と画像認識を行う前に既にかなり精度が高くなっている．これは，下位カテゴリーの長いキーワードをツイートする場合は，その食事画像である可能性が高いことを示している．一方，ラーメン，カレー，寿司は 7 割程度に留まっている．これに対して，食事検出エンジン，100 種類食事認識エンジンをそれぞれ単独で適用すると寿司を除いては 90%を上回った．しかしながら，寿司は 100 種類認識エンジンを適用しても 72.3%にしかなっていない．これは，元々寿司が桶に入っているものや，単品で多様なネタなどのために見た目のバリーションが大きく，認識対象として難しいためである．しかしながら，最終的に食事判別と 100 種類認識エンジンの両方を適用させると，92.7%まで向上させることが出来た．他の 6 種類については 99%台で，特にラーメン，チャーシュー麺については 300 枚中 1 枚のみの間違いでほぼパーフェクトであった．

最後に，位置情報付きツイートに含まれる食事画像を利用し，出現頻度が上位 1，2 位であるラーメンとカレーに関して，その地理的な分布の比較を行った．結果を図 2 に示す．左は 2 年 4ヶ月分，中央は 2012 年 12 月，右は 2013 年 8 月の地域毎の主要食事画像ツイートの分布を示す．赤印，黄印はそれぞれラーメン，カレーが最も多かった地域を示す．また，青印はお好み焼きの方がラーメン，カレーよりも上回った地域を示す．関西地方を期待していたが実際には広島周辺で，そのような地域が見られた．また 2012 年 12 月と 2013 年 8 月の比較により，全国的に冬はラーメンが多く，夏はカレーが多いことが分かる．またさらに，時間による違いの分析も行ったが，こちらは顕著

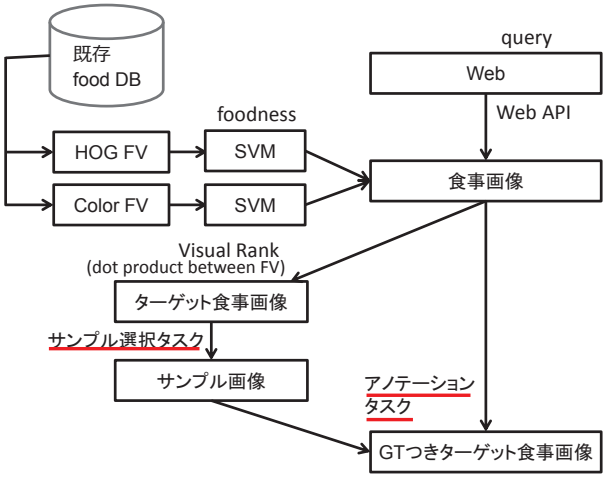

図 3　データセット自動構築の流れ

な違いは見られなかった．地域や時間に偏りのある食事の自動発見などもこうした分析の延長として興味深い．

以上，本節では，Twitter 食事画像を最新の一般物体認識手法による食事画像認識エンジンで認識し分析する試みについて述べた．ここでは既に収集した画像ツイートに対して分析を行った結果を述べたが，同時に Twitter stream API に食事キーワードを設定し，食事関連画像のみを収集する試みも行っている．その結果，1ヶ月で 2 万枚のラーメン画像を収集することに成功した．今後も収集を続け，4 年後には 100 万枚ラーメン画像データセットを完成させる予定である．また，今回，収集した大量の食事画像を fine-grained object recognition の手法を用いてさらに細かく分類することも予定している．これにより，例えば，ラーメン画像をネギ，チャーシューありの豚骨醤油などと細かく分類することを実現したいと考えている．

5. クラウドソーシングによるデータセット自動構築

本節では，既存の食事画像データセットに新たな食事カテゴリの画像群を追加する問題を考える．WebAPI を用いて収集した全ての画像をクラウドソーシングに用いるとコストが非常に高い．そのため，クラウドソーシングに用いる前に，画像のフィルタリングを行う．また，バウンディングボックス付与タスクには，ワーカーは提示する食事に対して無知であることが想像されるため，クラウドソーシングによりあらかじめサンプル画像を取得し，提示する．つまり，サンプルの選択と，バウンディングボックスの付与の 2 種類をクラウドソーシングする．サンプル選択タスクでは，食事性の高い上位の画像を Visual Rank でランキングし，その上位の画像を用いる．バウンディングボックス付与タスクには，食事性の高い画像を用いる．図 3 に，データセット自動構築の流れを示した．クラウドソーシングには Amazon Mechanical Turk (AMT) を用いた．

5.1 サンプル画像取得

データセットに入っていない食事カテゴリ 20 種類について Flickr(注2)，Twitter，bing(注3) より，収集した．海外で主に食べられるものについては，現地語および英語で収集した．収集した画像に対して，食事画像判別エンジンにより評価値を付与し，順序づけた．だが，食事性のみ評価を行っているため，対象カテゴリ以外の食事画像が含まれる恐れがある．クラウドソーシングによるサンプル選択タスクにおいても，一度に用いる画像数は少ない方が，ワーカーの仕事量が少なくなり，質の向上につながる．そこで，上位には対象の食事カテゴリの画像が相対的に多数存在するとし，それぞれ上位 300 枚を Visual Rank [13] により順序づける．Visual Rank に与える類似度行列には，HOG-FV，Color-FV 間の内積の平均の行列 S を用いる．L2 正規化された特徴ベクトル (FV) 間の内積は L2 距離と同等であり，FV 間の内積はよい類似度になっている [11, 14]．そして，以下で各要素を (0,1] に正規化し与え，次に Visual Rank 下での類似度行列の正規化を行った．また，[13] と同様に上位 m 枚にバイアスを設定した (m=100)．

$$S' = \frac{1}{S_{max} - S_{min}} S + 1 \quad (1)$$

$$S = \frac{1}{S'_{max}} S' \quad (2)$$

ここで，S_{max}，S_{min} はそれぞれ行列 S の最大，最小要素である．なお，FV に用いる GMM は，混合行列を求める際に作成した GMM を用いた．

Visual Rank での上位 30 枚をクラウドソーシングにより，サンプル画像を選択した．サンプル画像を選択するタスクでは，30 枚の画像を提示し，最大 7 つ対象の食事かつ一般的なものを選択する．画像検索サイトのリンクを設置し，そのリンクをクリックしない，またはサンプル画像を未選択であればタスクを提出できないようにした．ワーカーは画像検索サイトより，対象の食事の知識を取り入れることになり，質の向上になる．実験では，1 つのサンプル画像選択タスクに対して 6 セントを 4 人のワーカーに依頼した．最終的に，サンプルとして選択された数が多いものからサンプル画像として 5 枚選出した．

5.2 バウンディングボックス取得

ここでは，5.1 節により取得したサンプル画像を用い，バウンディングボックス付与作業，および対象の食事画像以外の画像の除去を依頼する．サンプル画像を提示，また，サンプル選択タスクと同様に，画像検索サイトへのリンクを設置した．だが，必ずしもリンク先を訪れる必要はない．画像 10 枚を表示し各画像に対して，対象の食事画像であれば，その食品の周り，かつなるべく背景を含まないようにバウンディングボックスを描画，対象の食事が存在しなければ存在しないのチェックをするように依頼した．質を向上させるために，タスクの提出時にバウンディングボックスの描画，存在しないのチェックを 3 枚以上の画像におこなっていない場合はその旨を表示し，提出できないようにした．また，画像検索サイトのリンクや，バウンディングボックスを描画した数，存在しない数はそれぞれカウンターを設置し，表示するようにしてある．さらに，バウンディングボックスが非常に小さい（表示している画像サイズの 3%以下），幅か高さが非常に小さい（表示している画像の幅か高さの 10%以下）場合は，描画した時点で消すようにした．付与した結果を評価，クリアは，タスクを回収した後に行うこともできるが，タスク進行中に行うことによって，ノイズの代わりに新たな回答を得ることができ，またワーカーの心理面においても監視しない場合と比較すると有用であると考える．

次に，ワーカーから回収した結果から最終的な判断を下す．ワーカーは必ずしも正確に依頼したタスクを実行してくれない，中にはタスクをわざと間違えるようなワーカーもいる．そのため，クラウドソーシングにより得た結果は，信頼できない場合があり，結果が妥当であるか判定する必要がある．ここで，画像中にバウンディングボックスを描画することは，自由度が

(注2)：http://flickr.com
(注3)：http://bing.com

図 4　収集対象の 20 種類の食事のサンプル

高いことに注目し，複数のバウンディングボックス間の分散が一定値未満のものを採用する．1 つのタスクを複数のワーカーに依頼したとき，複数のスパマーが描画した誤ったバウンディング間の分散が小さくなることを除き，極めて簡単な方法によりスパマーの付与したバウンディングボックスの結果は無視される．

具体的には，N 人のワーカーに対して（N-1）からなる組み合わせを考え，その中で 2 つ以上のバウンディングボックスの左上の座標，幅，高さが全て，表示した画像の X%未満に存在する場合，それらのバウンディングボックスを候補とする．最終的に，全ての組み合わせについて検討し，候補がない場合には，その画像は対象の食品でないとする．候補がある場合には，その中で最もバウンディングボックス間の分散が小さくなる組み合わせを選択する．そして，そのバウンディングボックスの組み合わせの平均を最終的なバウンディングボックスとして，その画像に付与する．ここで，X が許容誤差になる．この値を変更することで，質と取得できる画像枚数の調整をすることができる．本稿では，全てのカテゴリで固定し 15 とした．

ただし，ワーカーの半数が，対象の食事は存在しないとしていた場合は，その画像は対象の食品でない．また，1 人のワーカーのみがバウンディングボックスを描画した場合も，複数バウンディングボックスが得られなければ評価できないため，その画像は対象の食品でないとする．これにより，目的の食事カテゴリ以外の画像の除去を行う．

クラウドソーシングに用いる画像は，食事性が高いと評価された画像上位 300 枚から，ランダムに選択した．実験では，10 枚の画像を 1 タスクとし 5 セントを 4 人のワーカーに依頼した．

6. 実　　験

本節では，20 種類の新たな食事画像に対して，画像を収集し，クラウドソーシングによりバウンディングボックスを付与し，小規模なデータセットの構築を行う．既存食事画像データセットの食事カテゴリは，表 1 に示したとおりである．図 4 に，追加対象の食事のサンプルを示した．以前のデータセットには，一般にデザート，菓子類に分類される食品が「たいやき」しかなかった．今回は，デザートに分類される食品も複数含まれている．

6.1　未知カテゴリに対する食事画像判別エンジン

既存食事画像データセットで学習した食事画像判別エンジンを用いて，未知カテゴリの食事の判別が可能であるか評価を行う．まず，既存食事データセットに入っていない食事 15 種類を WebAPI を用いて 1000 枚以上収集した．

それらに対して，食事画像判別エンジンの評価値により順序づけたところ，上位 300 枚に 98.8%，上位 500 枚に 97.0%が食事が含まれていた．一方，下位 300 枚は 15.2%が食事画像であった．従って，効果的に食事画像を判定することができている．特に，上位の画像は，中央に大きく食事が写っている例が多く見られた．図 5 に，収集した画像のサンプルと，上位 300 枚にその食事性を付与した数が最も多い識別器に属する一部を示した．マンゴープリンは，卵焼きやコンスープの色，また茶

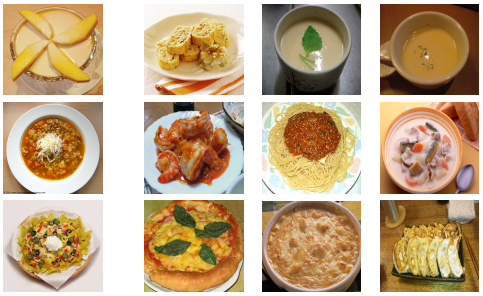

図 5　左の列は収集した画像のサンプル，右の 3 列はその食事性を付与した数の多い識別器に属する食品の一部

表 3　サンプル提示がタスクで有用であったか．それぞれ評価と回答の数を示している．

有用であった	まずまず	有用でなかった
763	103	22

碗蒸し，コンスープの形と似ていることから，うまく判別できたと考えられる．ナチョスは単一色でなく，様々な食品が含まれているため，難しいと考えられるが，ピザや餃子，グラタンなどと全体としてみると似ていないが，部分的に類似している部分があったため，うまく判別できたと考えられる．

ここで，FV のための GMM は，未知カテゴリの局所特徴量は考慮されていないが，高い精度を示した．未知カテゴリの画像においても，よい画像表現ができているためであると考えられる．そのため，既存の食事画像から食事性を学習し，未知の画像において食事かどうかを判定するアプローチは有効な手法であることが示唆される．

6.2　サンプル選択

そして，各カテゴリ上位 300 枚を Visual Rank により順序づけた結果，上位 30 枚に対象のカテゴリを含む割合は，80.9%から，88.6%に上昇した．また，上位 30 枚を確認したところ，画像の質もよくなっている印象を受けた．この 30 枚の画像をクラウドソーシングに用い，サンプルを選択する．

図 4 に示したサンプルは，本実験で行ったサンプル選択のタスクにおいて，最もスコアが高かった画像である．回収した 20 種類各 5 枚，計 100 枚の適合率は 98%であった．間違えはいずれも冷麺であった．麺ではあるものの，冷麺ではないと思われる画像が選択されていた．また，ちゃんぽんやナチョスなど食品の一部が写っていない画像も見られた．

取得したサンプルを用い，バウンディングボックス付与のタスクの精度，特に対象でない食品除去の精度を向上させることが目的である．バウンディングボックス付与タスクにおいて，サンプル画像がタスクで有用であったかのアンケートをした．表 3 は，それぞれ評価と回答のあった回数を示している．有用であったが約 86%と，有用であったことがわかる．

6.3　バウンディングボックス付与

合計 3730 枚の画像をクラウドソーシングに用い，データセットに追加された画像は 2706 枚，72.5%の回収率であった．次に，最終的にバウンディングボックスが付与された画像が，その食事画像であるかを調べた．その食品でない画像は，2706 枚中 74 枚であった．その内 33 枚は冷麺であった．だが，バウンディングボックスが付与された食事画像の中で 438 枚は 3 割以上食品の一部が写っていなかった．

バウンディングボックスの付与自体は，わずかに大きく付与されたものもあるが，概ね正しく食品の周りに付与されていることが確認された．しかし，得られたバウンディングボックス

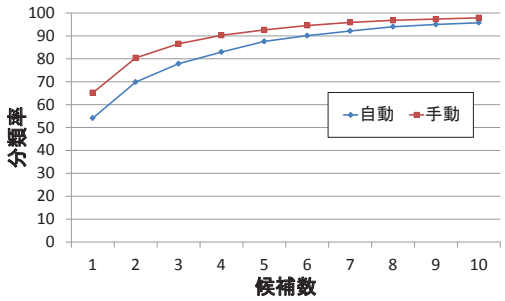

図 6　手動で構築されたデータセットと自動で構築されたデータセットでの分類率

の細かい評価は行っていない．今後，手動でつけたバウンディングボックスと比較を行うことで精度の評価を行いたいと考えている．

6.4 認識精度

クラウドソーシングにより構築したデータセットの評価を行う．比較として，既存食事画像データセットから 20 種類選択した．ここでは，単純にデータセットの比較を行うために，局所特徴量は 128 次元の HOG のみを用いた．局所特徴量は 3 スケール 4 ピクセルごとの dense sampling で抽出し，PCA で 64 次元に圧縮した．GMM のコンポーネント数は 64 とし，SPM は適用しなかった．各カテゴリ 20 枚ずつを評価画像としてランダムに選択し，評価することを 5 回繰り返した．分類率の平均を図 6 示す．手動で作成したデータセットは 65.2%，自動で作成されたデータセットは 54.1%の認識率であった．認識率の差は 11.1%であり，大きい差があることがわかる．

6.5 考察

クラウドソーシングに用いる画像について，Visual Rank により順序づけたとき，上位の画像の多くは，画像全体に写っていた．バウンディングボックスなしでも，学習データとして使用することができる．そこで，対象の未知カテゴリの知識を Adaptive SVM [15] などにより取り入れることで，さらなるフィルタリングが行え，効率よくなるように思える．

次に，認識精度が悪かった理由は主に誤った食事へのバウンディングボックス付与，また，収集した画像の質の悪さ，であると考えている．結果を集約する際に，誤った食事へのバウンディングボックスを完全に取り切れていない．条件を厳しくすることもできるが，そうせずに，バウンディングボックス付与作業の前，もしくは最終的な結果に対してクラウドソーシングによりフィルタリングを行うことで解決を図ることを考えている．現在は，バウンディングボックス付与と，対象の食品であるかの判定を一つのタスクで行っているために，一度に行う仕事量が多いおそれがある．そこで，仕事を分散させることで，質を向上させたい．収集した画像の質の悪さについて，同様にバウンディングボックス付与作業の際に，取り除くように指示はしているものの，今回の場合はあまり効果はなかった．さらに，手動で構築したデータセットと異なり，Web から収集した画像をそのままクラウドソーシングに用いているため，食品の全体が写っていない画像も多々であった．例として，図 7 は，認識が難しいと思われる画像であり，画像データセットに加えるべきかを悩む画像でもある．赤いバウンディングボックスは，最終的に付与されたバウンディングボックスを示している．バウンディングボックスは，特に問題なく付与されている．だが，金魚型マンゴープリン，アップルパイは，通常想像するマンゴープリン，アップルパイとは異なるように思える．同様

図 7　認識が難しい画像例

の金魚型マンゴープリンはデータセットに 114 枚中 7 枚含まれていた．ワッフルやベーグル，杏仁豆腐，釜飯も表面に他の食品が置かれている場合が多く難しいように思える．4. 節にも述べた通り，個性的な食事がツイート，アップロードされやすく，それが全体的な認識率の低下を招いたと考えられる．串カツに関しては，バウンディングボックスは間違っていないものの，くしの部分が含まれている．くしの部分をバウンディングボックスに含めると，背景が多く存在するという問題がある．だが，くしを含めないと串カツといえるのかという問題もある．食事の幅の広さと，それに応じた認識手法について検討する必要があるだろう．

文　献

[1] T Sakaki, M Okazaki, and Y Matsuo. Earthquake shakes twitter users: real-time event detection by social sensors. In *Proc. of the International World Wide Web Conference*, 2010.

[2] T. Kaneko and K. Yanai. Visual event mining from geo-tweet photos. In *Proc. of IEEE ICME Workshop on Social Multimedia Research (SMMR)*, 2013.

[3] Y. Kawano and K. Yanai. Real-time mobile food recognition system. In *Proc. of IEEE CVPR International Workshop on Mobile Vision (IWMV)*, 2013.

[4] 河野憲之, 柳井啓司. Foodcam: スマートフォン上でのリアルタイム食事画像認識による食事記録アプリケーション. 電子情報通信学会 データ工学研究会（DE）, 2013.

[5] K. Yanai. World seer: A realtime geo-tweet photo mapping system. In *Proc. of ACM International Conference on Multimedia Retrieval*, 2012.

[6] S. Vijayanarasimhan and K. Grauman. Large-scale live active learning: Training object detectors with crawled data and crowds. In *Proc. of IEEE Computer Vision and Pattern Recognition*, 2011.

[7] 大澤翔吾, 柳井啓司. クラウドソーシングによる食事画像認識モデルの自動構築. 第 5 回データ工学と情報マネジメントに関するフォーラム（DEIM）, 2013.

[8] V. C. Raykar and S. Yu. Ranking annotators for crowdsourced labeling tasks. In *Advances in Neural Information Processing Systems*, 2011.

[9] N. Dalal and B. Triggs. Histograms of oriented gradients for human detection. In *Proc. of IEEE Computer Vision and Pattern Recognition*, 2005.

[10] F. Perronnin and C. Dance. Fisher kernels on visual vocabularies for image categorization. In *Proc. of IEEE Computer Vision and Pattern Recognition*, 2007.

[11] F. Perronnin, J. Sánchez, and T. Mensink. Improving the fisher kernel for large-scale image classification. In *Proc. of European Conference on Computer Vision*, 2010.

[12] S. Lazebnik, C. Schmid, and J. Ponce. Beyond bags of features: Spatial pyramid matching for recognizing natural scene categories. In *Proc. of IEEE Computer Vision and Pattern Recognition*, 2006.

[13] Y. Jing and S. Baluja. Visualrank: Applying pagerank to large-scale image search. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2008.

[14] F. Perronnin, Y. Liu, J. Sánchez, and H. Poirier. Large-scale image retrieval with compressed fisher vectors. In *Proc. of IEEE Computer Vision and Pattern Recognition*, 2010.

[15] J. Yang and A. G. Yan, R.and Hauptmann. Cross-domain video concept detection using adaptive svms. In *Proc. of ACM International Conference Multimedia*, 2007.