1J4-OS-18a-7

クラウドソーシングによる食事画像データセットの自動構築

河野 憲之*1　　柳井 啓司*2
Yoshiyuki Kawano　　Keiji Yanai

*1*2電気通信大学大学院 情報理工学研究科
Department of Informatics, The University of Electro-Communications

本稿では、食事画像認識システムの認識対象を増やすために、自動で食事画像データセットの構築を行う。Web から収集した画像に対して、対象の食事画像であるかを判定し、それらの画像群に対して、クラウドソーシングを用いることで、デー タセットの自動構築を行う。実験では、100 種類の食事に対して、クラウドソーシングに用いる画像とデータセットの性能、一つのタスクでの仕事量とデータセットの性能について評価した。また、実際に食事画像データセット の構築を行い、高精度にデータセットが構築できることを示した。

1. はじめに

近年、モバイルデバイスを使って日々の食事記録をとるサービスが流行している。食事記録をとることによって、ユーザは自分の食習慣を確認することができ、ダイエットや栄養不良などの病気を防ぐことに有効である。食事記録をとる際に、テキスト入力や選択により食事記録をとる方法が一般的であるが手間が多く、継続した利用は難しい。そこで、画像認識によって、食事記録をとることを目標とした研究が行われるようになった [Yang 10, Chen 12, Matsuda 12]。これらの研究では、認識対象は 101 種類以下の小規模なシステムになっている。実際に食事は多数存在し、実用的な食事認識システムを構築することを考えた場合、より認識対象の食事を増やすことは有効である。

また、大規模で自動拡張されるデータセットには ImageNet データセット [Deng 09] がある。大規模なデータセットを手動で構築することは困難であるため、クラウドソーシングを用い、自動でアノテーションされている。

そこで本稿では、より実用的な食事画像認識システムの構築に向けて食事画像データセットの自動拡張を行う。さらに、様々の国の食事に対して収集対象とする。クラウドソーシングでは、ワーカーは様々な国の食事について未知であることが考えられるため、一つ目のタスクでその食事の一般的なサンプル画像を取得、二つ目のタスクでノイズ画像の除去、三つ目のタスクでバウンディングボックスを付与することで質の高いデータセットを自動拡張する。

そして実験により以下の評価を行う。

- サンプルを提示することの有効性
- 性能が完璧でないクラウドソーシングにおいて、タスクに用いる画像の重要性
- 一つのタスクの仕事量と構築されるデータセットの性能の関係

2. データセット自動構築の流れ

本稿では、Web からキーワードで収集した食事画像集合に対して、食事画像判別器によりノイズ画像の除去を行う。次に、クラウドソーシングにより 3 ステップでデータセットの自動構築を行う。データセット自動構築の流れを図 1 と以下に示した。

1. Web から収集対象の食事画像を収集する。
2. 収集した画像に対して、食事画像判別器により、食事らしさのスコアを付与する。
3. 転移学習を用いて食事画像判別器を再構築する。
4. 再構築した食事画像判別器で収集した画像の再評価を行う。
5. 食事画像判別器の評価値上位 30 枚をサンプル画像選択タスクに用い、対象の食事の一般的なサンプル画像を得る。
6. 食事画像判別器の評価値上位の画像をノイズ除去タスクに用い、対象の食事画像でない、またノイジーな画像を除去する。
7. ノイズ除去タスクで対象の食事と判定された画像をバウンディングボックス付与タスクに用い、バウンディングボックス付きの対象の食事画像を取得する。
8. バウンディングボックスが付与された食事画像をデータセットに追加する。

2.1 食事画像判別器

Web からキーワードで収集した画像に対して、そのキーワードの食事であるかを判別する識別器を構築する。識別器は食事画像認識の研究で用いられる [Kawano 13] に類似した手法に、転移学習 [Yang 07] を用い、さらに高性能にした。そして、構築された識別器によって画像を再評価する。

2.2 クラウドソーシング

収集、選別した食事画像をクラウドソーシングによって、GroundTruth としてバウンディングボックスを付与する。特に、ワーカーは様々な国の食事について知らないことが想定される。そのため、初めにその画像の一般的なサンプル画像を取得し、以降のタスクではそのサンプル画像を提示することでアノテーションの質を向上させる。クラウドソーシングは以下の 3 つのタスクから構成される。

連絡先: 河野 憲之，電気通信大学大学院 情報理工学研究科，
kawano-y@mm.inf.uec.ac.jp

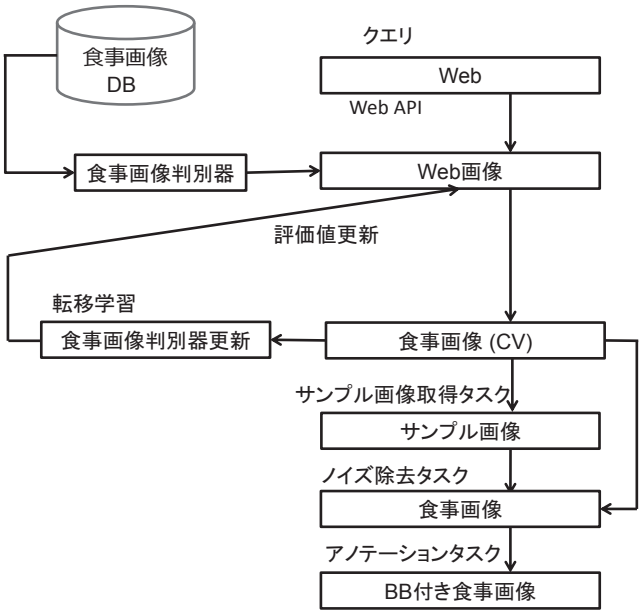

図 1: データセット自動構築の流れ

2.2.1 サンプル画像選択タスク

サンプル画像選択タスクでは、食事画像判別器の評価値上位 30 枚の画像をクラウドソーシングに用い、一般的なサンプル画像を最大 10 枚選択させた。ワーカーは対象の食事について未知である可能性があるため、画像サイトへのリンクを設置しそのリンク先に移動しないと HIT を提出できないようにした。他のタスクでもリンクは設置するが、リンク先に移動する必要はない。複数のワーカーからの結果の結合は多数決とし、複数票得た投票の多い画像から各カテゴリ 5 枚から 7 枚のサンプル画像を取得した。実験では、1HIT0.05 ドル、各 HIT5 人のワーカーに依頼した。

2.2.2 ノイズ除去タスク

ノイズ除去タスクでは、食事画像識別器の評価値上位の画像群から 25 枚をランダムに選択し、クラウドソーシングに用い、対象の食事画像であるかそうでないかをチェックさせた。未チェックが 5 つ以上あるとき、HIT は提出できないようにした。複数のワーカーからの結果の結合は多数決とした。実験では、1HIT0.03 ドル、各 HIT5 人のワーカーに依頼した。

2.2.3 バウンディングボックス付与+ノイズ除去タスク

バウンディングボックス (BB) 付与+ノイズ除去タスクでは、ノイズ除去タスクによって、ノイズでないと判定された画像 10 枚を選択し、クラウドソーシングに用い、対象の食事画像であれば、その食事の周りにバウンディングボックスを付与させた。そうでない場合は、対象でないチェックをさせた。バウンディングボックスは、食事の周辺になるべく背景を含まないように付与、皿があるときはなるべく皿を含まないように付与させた。また、バウンディングボックスが付与されたとき、それが小さすぎる場合 (バウンディングボックスの幅か高さが 1 割以下、面積が 3%以下) は、その時点で消去した。複数のワーカーからの結果の結合は、半数以上がノイズと判定していない、かつ複数のバウンディングボックスの始点と終点が x%以内 (x=15) に存在しているとき、それらのバウンディングボックスの平均を GroundTruth として付与した。実験では、1HIT0.05 ドル、各 HIT4 人のワーカーに依頼した。

表 1: 各タスクにおけるサンプル画像提示の評価 (割合)

	有用	普通	不要
ノイズ除去	89.59	7.90	2.52
BB 付与	91.68	7.02	1.31

いずれのタスクにおいても、現在のワーカーの進行状態を表示させた。例えば、画像に何らかの処理をした数、付与したバウンディングボックスの数などである。また、対象の食事画像であっても、ぶれている、物に隠れていて半分以上見えない、イラスト、パッケージ、結果に自信が持てない場合は対象の食事画像として扱わないように説明した。

3. 実験

実験では、100 種類の食事画像データセットの構築を行う。食事画像判別器の学習には、UECFOOD100[*1] データセットを使用した。UECFOOD100 データセットのサンプルを図 2 に、本稿で収集対象の 100 種類の食事のサンプルを図 3 にそれぞれ示した。

そして、サンプルを提示することの有効性と、性能が完璧でないクラウドソーシングにおいて、タスクに用いる画像が重要性、一つのタスクの仕事量と構築されるデータセットの性能の関係について評価を行う。

ここで、クラウドソーシングのそのタスクに用いた画像枚数に対して、ターカーが対象の食事画像であるとして回収された画像枚数の割合を回収率と定義する。また、適合率は画像群中にラベルが正しく付与された食事画像が含まれる割合を表す。

3.1 サンプル選択タスク

サンプル画像選択タスクは、全てのカテゴリで 1HIT でサンプル画像を取得できた。また、その適合率は 100%であった。

次に、サンプルを提示することによって、ワーカーにサンプル画像が有用であったか、有用でなかったか、どちらでもないの 3 段階で各 HIT ごとに自由回答形式で質問した。

サンプル画像提示の有用性について、ノイズ除去タスクでは 3495、バウンディングボックス付与タスクでは 5359 回答が得られた。表 1 に、得られた回答で有用、普通、不要それぞれの割合を示した。ノイズ除去タスク、バウンディングボックス付与タスクともに 90%程度サンプル画像が有用であるという解答を得た。また、不要と解答した割合は 3%未満と非常に小さく、サンプル画像は有用であることが示された。

3.2 タスクに用いる画像、タスクの仕事量とデータセットの評価

食事画像判別器

図 1 において、ノイズ除去専用のタスクは設定しない場合である。つまり、バウンディングボックス付与タスクでノイズ除去も行うため、タスクの種類は少ないが、一つのタスクにおける仕事量が多い構造になっている。また、転移学習による識別器の再学習も行わない。

食事画像判別器+転移学習

食事画像判別器を転移学習を用いて再構築した場合である。食事画像判別器と比較し、クラウドソーシングに用いる画像の精度が高くなっている。

*1 http://www.foodcam.mobi/dataset.html

図 2: 食事画像判別器の学習に用いた 100 種類の食事のサンプル

表 2: 各方法における、回収率と 100 枚の食事画像を得るための平均コスト (ドル)

	ノイズ除去		BB 付与		総量
	回収率	コスト	回収率	コスト	コスト
食事画像判別器	-	-	64.2	3.11	3.11
+転移学習	-	-	74.7	2.68	2.68
+ノイズ除去	80.9	0.74	86.7	2.31	3.16

表 3: 3 つの方法によるデータセットに含まれる食事画像集合の適合率

	適合率	gain
食事画像判別器	91.10	-
+転移学習	94.19	+3.09
+ノイズ除去	97.83	+3.64

食事画像判別器+転移学習+ノイズ除去タスク

図 1 の流れである。食事画像判別器+転移学習とは、タスクの種類は増えるが、一つのタスクで行う仕事量が分散され少なくなっている点が異なる。

表 2 に、それぞれの方法における回収率と 100 枚の食事画像データセットを構築するために必要なコストを示した。食事画像判別器を再構築し、クラウドソーシングに用いる前にノイズの除去を行うことでより回収率が高く、14%低コストであることがわかる。さらに、ノイズ除去タスクを加えた場合、バウンディングボックス付与タスク自体の回収率は向上した。だが、ノイズ除去タスクに用いたコストを考慮すると、コストは少し多くかかることがわかる。我々の実験では、バウンディングボックスのタスクに取り組むワーカーの数は 4 人と少ない。回収率や性能を高めるためにより多くのワーカーに仕事を依頼すれば、コストの差はほとんどなくなると考えられる。例えば、[Vijayanarasimhan 11] では 10 人のワーカーにバウンディングボックス付与タスクを依頼している。

次に、表 3 に各方法で構築されたデータセットにおける食事画像集合の適合率を示した。食事画像判別器では、91.10%の適合率であった。一方、判別器の再構築により、画像の選別性能を向上させた場合、94.19%の適合率と 3.09%性能が向上した。表 2 から回収率も向上し、低コストで構築可能であることが示されただけでなく、データセットのノイズも減り、質も向上することがわかる。これは、クラウドソーシングでのアノテーション性能が完璧でなく、ノイズ画像が対象の食事画像であると判定され、バウンディングボックスが付与されることを防いだためである。さらに、ノイズ除去タスクを追加した場合、3.64%適合率が向上した。これは、タスクを分散することにより、仕事量が減る。また、2 段階で異なるワーカーによりノイズのチェックをさせることで、誤りを減少させることができたためだと考える。だが、ノイズ除去タスクを追加したことにより、バウンディングボックス付与タスクのコストは減ったが、全体では増えている。

以上より、クラウドソーシングに用いる画像のノイズを除去しておくことは、コストの削減のみでなく、データセットの質も向上することが示された。さらに、タスクをわけ、仕事量を分散することでデータセットの質が向上することが示された。そして、正しくラベル付けがされている食事画像の割合が 97.83%とノイズの少ないデータセットが構築可能であることを示した。

なお、本実験では GroundTruth として付与されたバウンディングボックスの詳しい評価は行っていない。だが、目視で確認したところバウンディングボックスの精度は高かった。バウンディングボックスを付与することは自由度が高いため、意図した通りに仕事をしないワーカーの結果は、複数のでたらめなワーカーが同じ位置にバウンディングボックスを付与した場合を除き無視されるため、バウンディングボックスの付与は正しく行えたと考える。実際、図 4 は、一人のワーカーが 1HIT10 枚に付与したバウンディングボックスであるが、無視

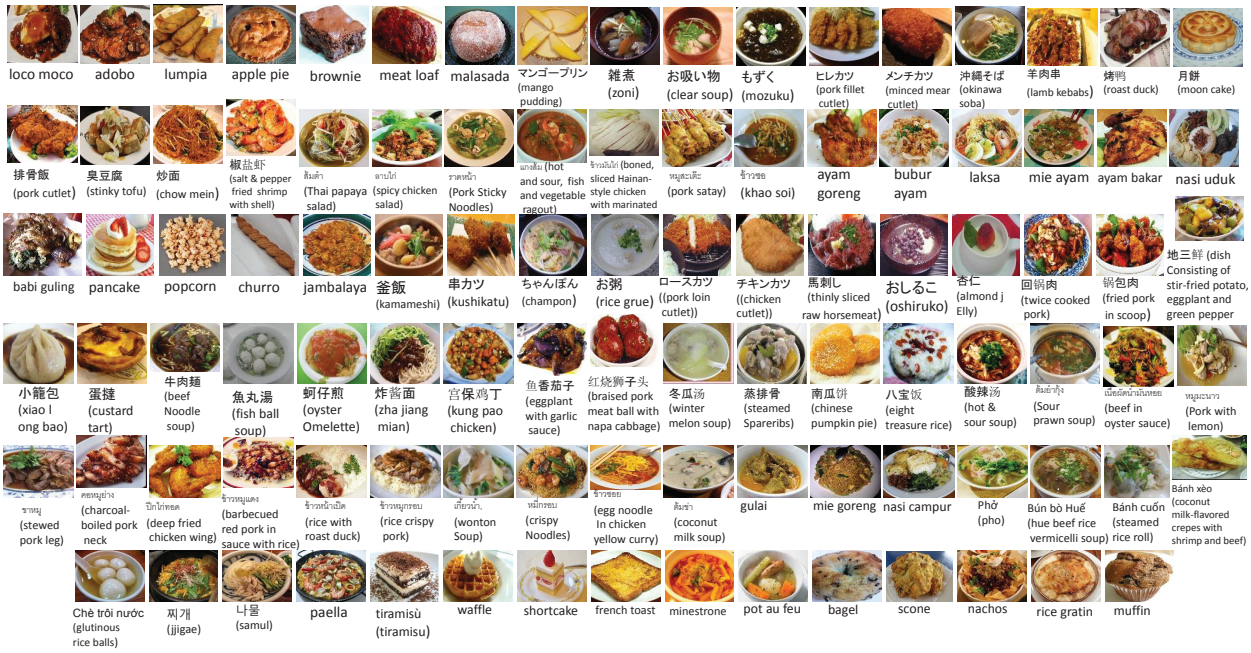

図 3: 本稿で収集対象の 100 種類の食事のサンプル

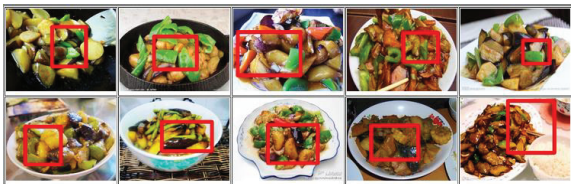

図 4: 意図した通りに仕事をしないワーカー

された。

4. 結論と今後

食事画像認識システムの認識対象を増やすために、自動で食事画像データセットの構築を行った。Web から収集した画像に対して、対象の食事画像であるかを判定し、それらの画像群に対して、クラウドソーシングを用いることで、データセットの自動構築を行う。実験では、100 種類の食事に対して、クラウドソーシングに用いる画像とデータセットの性能、一つのタスクでの仕事量とデータセットの性能について評価した。また、実際に食事画像データセット の構築を行い、高精度にデータセットが構築できることを示した。

今後は、既存の食事画像データセットと自動で構築した食事画像データセットとの質的な違いについて検討する。また、本論文では、収集対象を食事に限定したが、他のカテゴリについても実験を行う。

参考文献

[Chen 12] Chen, M., Yang, Y., Ho, C., Wang, S., Liu, S., Chang, E., Yeh, C., and Ouhyoung, M.: Automatic Chinese Food Identification and Quantity Estimation, in *SIGGRAPH Asia 2012 Technical Briefs* (2012)

[Deng 09] Deng, J., Dong, W., Socher, R., Li, L.-J., Li, K., and Fei-Fei, L.: ImageNet: A Large-Scale Hierarchical Image Database, in *Proc. of IEEE Computer Vision and Pattern Recognition* (2009)

[Kawano 13] Kawano, Y. and Yanai, K.: Rapid Mobile Object Recognition Using Fisher Vector, in *Proc. of Asian Conference on Pattern Recognition* (2013)

[Matsuda 12] Matsuda, Y. and Yanai, K.: Multiple-Food Recognition Considering Co-occurrence Employing Manifold Ranking, in *Proc. of IAPR International Conference on Pattern Recognition* (2012)

[Vijayanarasimhan 11] Vijayanarasimhan, S. and Grauman, K.: Large-scale live active learning: Training object detectors with crawled data and crowds, in *Proc. of IEEE Computer Vision and Pattern Recognition*, pp. 1449–1456 (2011)

[Yang 07] Yang, J. and Yan, A. G., R.and Hauptmann: Cross-domain video concept detection using adaptive svms, in *Proc. of ACM International Conference Multimedia* (2007)

[Yang 10] Yang, S., Chen, M., Pomerleau, D., and Sukthankar, R.: Food recognition using statistics of pairwise local features, in *Proc. of IEEE Computer Vision and Pattern Recognition* (2010)