Vol.2020-MPS-127 No.15
2020/3/2

深層学習によるインスタグラム画像からの流行抽出

西田　奈生†　　金本　玲花†　　松本　尚†

概要：近年，深層学習を使用したデータ分析・予測が進んでいる．Twitter や Instagram 上の文字情報や写真から多くの情報を得ることができ，それらの情報価値はとても高く，企業のマーケティングツールとしても使用されている．Instagram に投稿されている大量の画像には投稿者の嗜好が反映されていると考えて，深層学習を用いて分析することにより，世の中の流行を割り出せると考えた．本研究では，この考え方に基づく流行抽出方法の提案と具体的な実装方法の検討を行う．本稿では，Instagram に投稿されている，ある期間，ある場所において投稿された画像を収集し，その中から流行を割り出したい対象物を切り出して，その対象物をさらに詳細に分析することにより，対象物の中での「流行」を判定するための手法について述べる．

Trend extraction from Instagram images by deep learning

NAO NISHIDA†　　REIKA KANEMOTO†　　TAKASHI MATSUMOTO†

1. はじめに

Instagram [1]は，日常の瞬間をとらえた写真や動画を簡単にシェアできる SNS である．一日当たりのアクティブユーザー数 5 億人以上であり，今話題，流行となっている物体を含む画像や動画が毎日大量に投稿されている．それ故情報価値がとても高く，企業のマーケティングツールとしても使われている．実際に，同じ SNS である Twitter 上の文字情報から AI によって事件・災害などの発生をいち早く探知しマスコミに情報を提供する「株式会社 JX 通信社」 [2]がある．本研究では，文字情報ではなく Instagram に投稿されている，ある期間，ある場所の大量の画像情報を収集し，その中から流行を割り出したい対象物を切り出して，その対象物をさらに詳細に分析することにより，対象物の中での「流行」を判定する．本稿では，流行を抽出するための手法の提案と，流行抽出に必要とされる物体切り出し器や物体認識器について説明する．

本稿の構成は以下の通りである.2 章で流行抽出手法の提案を行い，3 章で画像収集の方法について説明し,4 章で使用する深層学習技術について述べる．5 章で物体検出の新規学習についての説明を行う．

2. 流行抽出手法の提案

まず Instagram からの画像を web スクレイピングにより収集する．詳しくは 3 章で説明する．収集した画像を CNN（Convolutional Neural Network）による物体切り出し（物体検出）器にかけ，流行を判定したい物体を切り出し，切り出し画像を生成する．そして，それらの画像をさらに詳細に分析することで流行を割り出す．詳細な分析の一例としては，CNN による詳細分類器にかけ，切り出した物体中の種類ごと写真への頻出度を調べ，一番多く分類された種類（クラス）が最も「流行」している種類と推定される．他の詳細な分析としては，切り出した物体の色の分布を調べて，何色の物体が流行っているかを検討するということ等も考えられる．Instagram 画像の収集を時間の隔たった期間を対象に行えば，流行の移り変わりの分析も可能になる．画像の撮影場所を変化させることにより，地域による流行の差異の分析も可能である．

物体の切り出しには YOLO を用いた．YOLO とは一枚の画像内の複数の物体の切り出しと識別が可能な CNN である．YOLOv1 から始まり，v2，9000，v3 があるが，本研究では YOLOv3 と YOLOv1 を主に使用する．これらについても詳しくは 4 章で説明する．本稿で示す最初の流行抽出例では，切り出した物体をさらに詳細に分類することで，その物体内での流行を探ることにする．そのための詳細分類器の学習において，学習画像データ，テスト用データは Google 検索サイトよりスクレイピングを行い収集する．そして CNN の認識器として VGG-16 を用いて，転移学習による Fine-tuning を行う．この詳細分類器によって各クラスに分類された画像の枚数をカウントし，カウント数が一番大きいクラスを最も「流行」しているクラスとする．以下流行抽出の流れを図示する（図 1）．

† 奈良女子大学
Nara Women's University

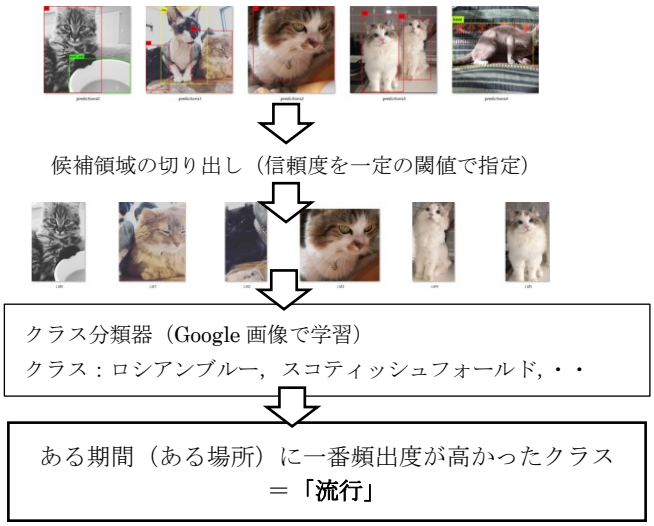

図 1：流行抽出の流れ

3. 画像収集

Instagram の画像，そして分類器で使用する学習用データ画像はインターネットから全てスクレイピングを用いて収集する．（ただし，2019 年 8 月ごろから撮影場所を特定した画像の自動収集は規約上できない状態にあるため，それ以降は撮影場所を特定した画像に関しては手動で収集している．）

スクレイピングとはウェブサイトから収集者の意図に基づいて情報を自動抽出する行為である．様々な方法があるが，今回は HTML 取得に Selenium や Requests という Python ライブラリを，HTML の解析・抽出に Beautiful Soup4 というライブラリを用いて行った．これによって投稿された日付情報がある HTML の要素へアクセスし，日付情報を抽出する．対象の web サイトには JavaScript を使用しているものがあり，これによって HTML を取得する方法が変わってくる．Instagram では JavaScript が埋め込まれているため，JavaScript 実行後の HTML を Selenium で取得した．また Selenium 自体でも HTML を解析・抽出できるが，今回解析・抽出には Beautiful Soup4 を用いる．Yahoo，Google では今回はそのままの HTML で可能なため Requests によって取得する．実装環境は Python 3.6.3 であり，ブラウザは Google Chrome を用いた．

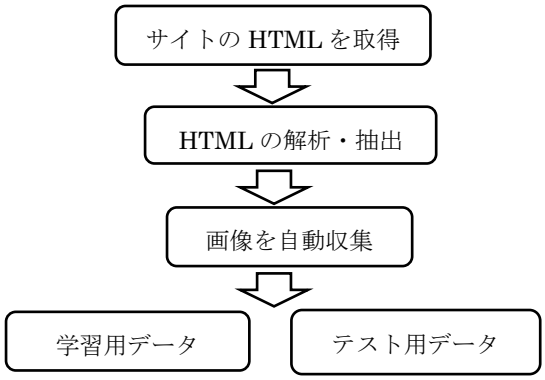

図 2：スクレイピングの流れ

4. 使用する深層学習技術

4.1 YOLO [3]

流行を抽出するために流行を調べる物体を Instagram 画像から切り出す必要がある．物体を識別して切り出す作業を物体検出と呼ぶ．物体検出の深層学習を使った手法には様々な種類がある．本研究では他よりも高速であることが知られている YOLO を用いる．YOLO は，入力画像の中からオブジェクトらしい候補領域を多数切り出し，同時にその候補領域について，オブジェクトの各クラスの確率を推測し，大きい確率の候補領域を検出結果として出力するという CNN である．YOLOv1 から始まり，YOLOv2，YOLO9000，YOLOv3 がある．YOLOv2 からは完全結合層を無くし，畳み込み層だけで構成されている． YOLO9000 は 19 層で，画像データベース ImageNet [4]内の 9000 カテゴリーに属するオブジェクトの切り出しと，大まかなクラス分類が可能になった．YOLOv3 は 53 層で，3 つの異なる大きさのフィルターによって特徴マップを作り出すことで精度が向上されている．本研究では物体検出に YOLOv3 または YOLOv1 を用いる．

4.2 詳細分類器の作成

詳細分類器の CNN として VGG-16 を用いて，転移学習による Fine-tuning を行う．流行（本例では東京で注目されている猫の種類）を割り出すため，クラス分類器を作る．スクレイピングで取得した種類別の猫の Google 画像を用いて詳細分類器のデータセットを作成する．実装環境に用いたフレームワークは Theano0.8.2（Keras1.0.8 のバックエンド）である．

詳細分類機では ImageNet の定義に従って，猫 6 種の細分類を行うものとする． 6 種の定義は以下の通りである．

Angora cat：ペルシャ猫に似た長い毛を持つ猫
Burmese cat：シャム猫に似ているが，濃い茶色または灰色 の無地で短い毛を持つ猫
Persian cat：長い毛を持つ猫
Siamese cat：黒い耳の顔と尾の先端の淡く，細身の短い毛の 青い目の猫
Tabby cat：黒でまだらな灰色または黄褐色の毛を持つ猫
Tortoisesh cat：全体に黒っぽく，一部の毛が赤茶色や黄土色 の毛の猫

Angora cat に関しては補足として，くさび型の頭部に付け根の広い大きな耳はとがっており，色は白だけではない． 3 章で収集した各猫の種類ごとの Google 画像の枚数は以下のようになった（表 1）．これらの画像内には指定した猫の種類でない画像も混入している．またクラス分類器の学習では物体が中心にある質の良い画像が求

められるため，ここでも YOLOv3 を用いて，これらの画像から更に「猫」の物体の切り出しを行う（表 1）． 今回はホールドアウト検証を行う．ホールドアウト検証とは，機械学習におけるデータのテスト方法の 1 つであり，教師データをトレーニングデータセット，バリデーションデータセットに任意の割合で 2 分割し， 学習済みモデルの精度をバリデーションデータセットの誤差をみながら測定する手法である．今回はこのトレーニングデータセット，バリデーションデータセットにランダムに選んで振り分ける割合を 8：2 にし，テストデータは画像データセット ImageNet から各枚数ダウンロードした（表 2）． YOLOv3 にかけたことによる枚数の減少の補完と，分類器の精度向上のため，各データセットに対し，回転，拡縮，平行移動，せん断などの変換を行い，各データセットを 5 倍に拡張した．変換の詳細は，1 枚の画像に対し以下の 6 種全ての変換を施し新たな画像を生成する．ランダムな値で元画像から 5 回変換を行い 5 倍に拡張した．

拡縮：アスペクト比を固定し， 1 倍
回転の角度：-15~15 度以内
せん断：-20~20 度以内
平行移動：ピクセル-30~30 以内
反転
伸縮：アスペクト比を固定せず，縦横 1/1.3~1.3 倍以内

	元画像数	切り出し後
Angora cat	252	202
Burmese cat	368	213
Persian cat	319	186
Siamese cat	444	343
Tabby cat	337	315
Tortoisesh cat	427	261
合計	2147	1520

表 1：Google 検索で取得した全画像の枚数と切り出し後の画像枚数

	トレーニングデータセット	バリデーションデータセット	テストデータセット
Angora cat	160	42	77
Burmese cat	170	43	143
Persian cat	150	36	140
Siamese cat	270	73	139
Tabby cat	250	65	132
Tortoisesh cat	200	61	136
合計	1780	320	767

表 2：各データセットへの振り分け

4.3 VGG-16 [5]による転移学習

VGG-16 の学習済みモデルで転移学習による Fine-tuning を行った．畳み込み 13 層と結合層 3 層のニューラルネットワークに，ImageNet 内の 1000 カテゴリー，120 万枚を学習させている．Fine-tuning ではこの学習済みモデル VGG-16 の重みを分類器のモデルの初期値とし，再学習させた．
手順は以下の通りである．

①出力層を除いた 1 から 15 層までの構造を作成する．
②VGG-16 の重みを初期値としてセットする．
③出力層を 6 クラス分類用にする．

②において，VGG-16 の重みをダウンロードしておき，15 層目と 16 層目の間で VGG-16 の重みを読み込む．③においては，実際の Keras の実装では model.add(Dense(6, activation=’softmax’)を最後に加えることで出力層を 6 ユニットにしている．

バリデーションデータセットの誤差を見ながら最適なエポック数を求めておく．少ないエポック数で高い性能が得られることが示されており，今回は学習終了エポック数を 50 に指定した．ホールドアウト検証 1 回目ではエポック数 16 の時がバリデーションデータセットの正解率が一番高く，ホールドアウト検証 2 回目ではエポック数 48 の時が一番高くなった．推測の際にこのエポック数を指定した．バッチサイズ 32 とし，学習させ，テストデータに対する推測を行い，各正解率を求めた．（表 3）

	正解	不正解	計	正解率
Angora cat	62	15	77	80.5%
Burmese cat	55	88	143	38.5%
Persian cat	100	40	140	71.4%
Siamese cat	131	8	139	94.2%
Tabby cat	107	25	132	81%
Tortoisesh cat	118	18	136	86.8%
合計	573	194	767	74.7%

表 3：テストデータの正解率

Burmese cat（バーミーズ）では正解率が落ちるが，全体では 74.7 %と良い正解率となった．この分類器を用いて Instagram 画像から切り出した猫の画像の推測・分類を行う．

検出された 233 枚の画像を，YOLOv3 にかけ，信頼度 80%以上で切り出し，171 枚の新たな画像を生成した．それを作成した詳細分類器にかけクラス分類を行い，画像ごとに一番大きな推測値であったクラスに 1 ポイントとしてポイントを加算していき，クラスごとに算出した．結果は表 4 のようになった．

Angora cat	14
Burmese cat	8
Persian cat	12
Siamese cat	38
Tabby cat	69
Tortoisesh cat	30
合計	171

表 4：分類結果

6 つの種類の猫でない画像の場合も，6 つの内のどれかに分類されてしまい，結果の数値が大きくなってしまう欠点があるため，結果より，Tabby cat（トラネコ）が流行しているというよりは，Tabby cat あるいは Tabby cat に類似性がある猫が流行しているといえる．流行抽出における詳細分類器の信頼性を確かめるために，Angora cat と判定された 14 枚の画像を詳しく見てみる．図 3 は 14 枚の画像中の一枚である．目視で Angora cat と判断されたのは 4 枚，Google 検索結果の画像と照らし合わせて Angora cat と判断されたのは 4 枚であり，合計 8 枚より，この分類器が 57.1%(=8/14)程度の精度であることが分かる．この詳細分類器の信頼度は改善余地がまだまだ大きいが，この流行判定が大枠においては正しいことがわかる．

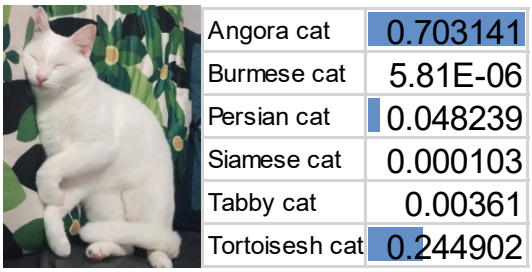

図 3：Angora cat に分類された画像の分類結果の一例

5. 物体検出の新規学習

東京で撮影された猫の詳細分類では，学習済みの YOLO を用いて Instagram 画像から猫の画像を切り出し，それを詳細分類器にかけた.しかし，この方法では学習済み YOLO の検出対象になっている物体に対してしか流行を調査できない．この問題を解決するために，検出対象となる新たな物体の写っている画像を収集し，YOLO に対して新規学習を行い，新たな物体を検出可能にする．新規学習では 4.1 節で説明した YOLOv1 を用いる．今回は，流行を抽出するという観点から「バッグ」を抽出するための物体検出器の作成を試みる．ネットで収集したバッグの画像に対して，BBox-Label-Tool を使用し，画像内にあるオブジェクトの範囲を長方形で囲み，座標を取得し，位置情報を作成する．BBox-Label-Tool とは，Python で稼働する画像の範囲指定ツールである．YOLO では学習実行時に教師情報として物体の識別情報の他に画像内の「オブジェクトの位置情報」としてのバウンディングボックス情報を必要とする．

5.1 Caltech101 [6]の画像での学習

まず，猫画像の詳細分類でも参考にした，書籍「実装ディープラーニング」 [7]を参考にして，Caltech101 の画像とインターネットからスクレイピングを用いて収集したバッグの画像を用いて 3 種類(airplane，motorbike，bag)で学習を行った．それぞれ 160 枚ずつ画像を用意した．学習回数は 8000 回とした．学習により作成された重みを用いて Pascal VOC [8]の画像のデータセット(aeroplane，motorbike) をテストデータとしてテストを行ったが，誤検出が多く，全く使い物にならなかった．理由として考えられることは，Caltech101 の画像の特徴が，物体の横向きの全体がはっきり写った画像であり，サイズも統一で物体 1 つだけの画像が多く，学習させたい物体が比較的画像の中心に位置しているため，様々な角度やシチュエーションで撮影された Pascal 画像とは相性が悪かったようである．（図 4）

図 4：Caltech101 の画像

Instagram 画像として掲載される写真が Caltech101 のような画像とは限らないため，Pascal VOC のデータセットのように，学習させたい物体以外も多く写っており，物体の位置やサイズもバラバラで、物体を様々な角度から写した画像（図 5）を用いて学習させる必要があると痛感した．このため，Pascal VOC 画像の aeroplane と motorbike の一部を学習データとして使用することに変更した．

図 5：Pascal VOC の画像

5.2　学習方針を変更して実験

当初の方針としては，YOLO に新規学習をさせて物体検出器として使用するために，物体の種類を徐々に増やして学習させる手法で実験を行っていた．しかし，この方法だと学習させていない物体に対する誤検出がなかなか減らないという問題があることが判明した．このため，検出させたい物体に対して学習させた重みを使って，学習させていない物体の画像データで推論を行い，その誤検出したデータを，検出する物体数を増やすことなく学習に追加し，追加学習を行う手法で学習を行うことにした．つまり，肯定情報のみの教師データから否定情報を含む教師データを使用することにした．YOLO は 1hot タイプの認識器ではないため，この方式がうまく機能すると予想した．

まず，実験的に Pascal VOC の 2 種類（aeroplane, motorbike）の画像を 300 枚用いて 20000 回学習を行った．学習時に 1000 ずつ作成された重みを用いて学習に使用しなかった aeroplane と motorbike の画像をテストデータとして推論を行った．

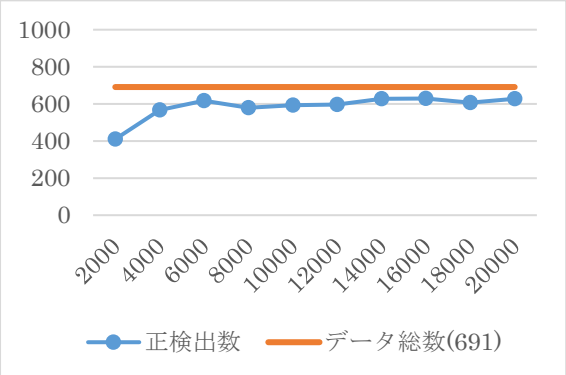

図 6：テストデータに対して正しく検出された数

同じテストデータに対して誤検出の数が一番少ない学習回数は，18000 回であった．2000 回では誤検出は少ないが，検出率も低いため，2000 回は除く．（図 6）また，18000 回の重みを使って検出対象になっていない Pascal VOC データセットの cat，dog の画像データを使って推論を行い，誤検出がどれくらいあるかを検証した．cat では，誤検出が 131 枚，dog は 261 枚であった．データから誤検出された画像データの半分の学習データを否定情報の教師データとして追加し，さらに 10000 回学習させた.また，bird もテストデータとして誤検出テストを行い，10000 回学習させた後に bird の誤検出画像を否定教師データとして追加し，さらに 10000 回追加学習を行った．

以上の結果から，複数種類の物体検出を行う物体検出器を生成しなくても，精度の高い物体検出器が構成できると判断した．このため，「バッグ」のみを抽出させるための新規学習を YOLO に対して行った．学習回数は 10000 回とし，画像数は 300 枚である．10000 回学習させたデータに対して同様にして cat，dog の画像データで推論を行い，誤検出がどれくらいあるか検証した．その後，その誤検出したデータを否定教師データとして学習に追加し，さらに 10000 回追加学習を行った．学習に使用したバッグの画像が 300 枚であるため，追加学習に使用した誤検出したデータもそれぞれ 300 枚追加した．

5.3　物体検出器の学習結果

Pascal VOC の画像で学習された重みは，複数の物体や画像の中心にない物体も検出している．また，物体がはっきり写っていない場合でも検出されている．

また，手法を変更し学習させたデータに検出対象になっていない物体の画像データで推論を行い，その誤検出したデータを学習に追加し，再学習を行う手法で学習を行った結果，誤検出の数が大幅に減少した．（図 7）

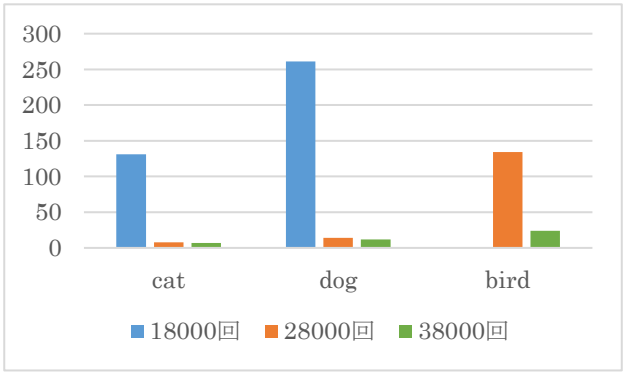

図 7：3 種物体検出の誤検出の変化

この結果から，誤検出した画像データを学習データに組み込み，再度追加学習させる手法が誤検出を減らすために有効であると考えられる．また，再度学習を行うことで対象物が写っている画像の検出率も多少上がることが分かった．

また，バッグ単体検出器の学習においては実験的に行った 2 種物体検出器の結果と同様，誤検出の数が大幅に減少した．（図 8）

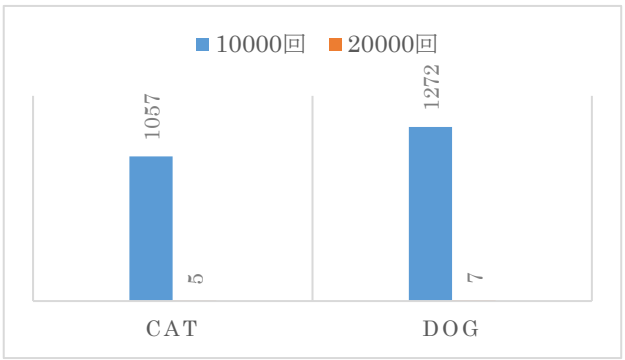

図 8：バッグ検出器の誤検出の変化

バッグ単体検出器の検出対象ではない物体の画像データを使った推論では，10000 回のときは，画像に写っているバッグ以外の物体にもバッグと推測していたが，

20000 回のときにはバッグだけを検出する数が増えており，無駄な推測が減っている．しかし，バッグの肯定的教師データとして使った画像がバッグだけがはっきり写っている画像が多いため，人がバッグをかけている画像等ではまだ誤検出が起きている．

6. まとめ

本稿では，流行を抽出するための手法の提案と，流行抽出に必要とされる物体検出器と詳細分類器について説明した．

まず，ウェブスクレイピングと深層学習の物体検出器を用いて，インターネットから流行を抽出する対象となる物体画像を切り出した．その切り出した画像をさらに分析することで流行を抽出する．この分析方法の一つとして，詳細分類器によるクラス分けを提案した．様々なウェブサイトから画像を取得し，より多くのクラス数，枚数の画像で学習を行わせることにより，様々な詳細分類器の作成が可能となる．

今回，詳細分類器の作成については猫の種類の分類を取り上げた．物体検出に関しては，猫は元々学習済みの YOLO で検出できるため，YOLO を学習させる必要は無く，詳細分類器として使った VGG-16 のみの学習を行った．しかし，この手法では YOLO で切り出し不可能な物体に関しては，流行抽出ができない．このため，YOLO に新規学習させる方法について検討を行った．当初は，検出可能な物体を徐々に増やしていく予定であったが，検出物体の種類を増やしていく手法より，検出したい物体を学習させた重みデータを使って，検出対象になっていない物体の画像データで推論を行い，誤検出した画像データを否定教師情報として学習に追加し，追加学習させる手法で行う方が，計算時間を節約して，検出対象ではない物体の画像への誤検出の数が大幅に減少させることが可能であり，物体検出の精度が上がることが分かった．

また，学習においてどのような画像を使用するかが検出精度に大きく関わることが実験結果から分かった．SNS など様々な写真から物体を検出させるためには学習させる物体が複数写っている画像や様々な角度から写されている画像のデータセットを使用することが必要と思われる．

今後の課題としては，高精度で物体を検出可能な物体検出器の作成方法を確立し，一般の人々が興味をもつような流行分析ツールを仕上げていくことである．また，詳細分類器に関しても，ハイパーパラメータの調整等でさらなる精度向上が必要である．また流行を抽出するという面で，SNS で多く扱われているアクセサリーやスイーツの検出を行うための物体検出器および詳細分類器を作成する．今回は YOLOv1 を使用して物体検出器の学習を行ったが，YOLOv3 を用いて学習させることでより物体検出器としての精度を上げることが可能であると考えられる．

参考文献

[1] “Instgram,” Available: https://www.instagram.com/.

[2] “株式会社 JX 通信社,” Available: https://jxpress.net/.

[3] S. D. G. A. F. Joseph Redmon, “You Only Look Once: Unified, Real-Time Object Detection,” 2016.

[4] “ImegeNet,” Available: http://www.image-net.org/.

[5] K. S. Zisserman, “VERY DEEP CONVOLUTIONAL NETWORKS FOR LARGE-SCALE IMAGE RECOGNITION,” 2015.

[6] “Caltech101,” Available: http://www.vision.caltech.edu/Image_Datasets/Caltech101/.

[7] 藤田一弥，高原歩，“実装ディープラーニング,” 平成 28 年.

[8] “Pascal VOC,” Available: http://host.robots.ox.ac.uk/pascal/VOC/.