

リアルタイムな動画内物体認識技術を用いた物探しシステム

— FineTop: ストレスフリーな物探し —

1. 背景

物探しは我々にとって非常に身近で重要な課題であると言える。我々は記憶やバイアスに頼って物探しを行っている上、物を見落とししてしまうことも多いため、物探しに多くの時間をかけてしまうことがしばしばあり、これはストレスフルな事象である。また例えば標識や警告などの物体は、見落とすことで取り返しのつかない重大な事態を引き起こすことにもなりかねない。人間に代わって物を探し、見落としを防止してくれるシステムの存在は、我々の生活をより豊かで安全なものにする力を秘めていると言える。

近年ではコンピュータが画像から物体を認識する技術は大いに向上してきてはいるものの、依然誰でも気軽に物探しを行えるようなシステムは存在していない。事前準備や環境の整備が必要で、誰もが気軽に使える技術ではなかったためである。そのような準備が必要なく、容易かつ即座に利用できるようになる物探しシステムを開発し、豊かで安全な社会の実現に貢献したいと考えた。

2. 目的

本プロジェクトの目的は、人間に代わってリアルタイムに物体を認識するシステムを提供し、ユーザが豊かで安全な生活を送れるようにすることである。このシステムを利用することで、ユーザはより簡単に物が見つけられるようになり物探しに対する煩わしさが軽減される他、標識などの見落とし防止に利用することでより安全な生活を送ることができるようになる。

本プロジェクトでは、物探しの中でも特に商品探しに焦点を絞り、ユーザの探したい商品をリアルタイムに探し出すウェブアプリケーションを開発する。

3. 開発の内容

3.1. アプリケーションの機能

ユーザが探したい商品(物体)を写した画像1枚を入力すると、カメラの映像にその商品が映った時にユーザにフィードバックを返すウェブアプリケーション“FineTop”を開発した。

使い方は次の通りである。

1. ユーザはまず、画面下部の検索窓にキーワードを入力し、探したい商品の画像を検索する。
2. 目的の商品画像を取得したら、モデルを読み込む。
3. モデルの読み込みが完了すると、アプリケーションは即座に目的の物体を認識できるようになる。カメラにその物体が写ると、アプリケーションは背景色を白からオレンジに変化させ、またバイブレーションすることで、ユーザに対し物体を認識したことをフィードバックする。

検索画像は Amazon の商品画像を利用したものであり、多くの商品について検索可能である。また検索して取得した画像のみならず、ユーザ自身で画像を用意しアップロードすることでも本システムを利用し物体を探すことも可能である。

図 1 の例では、システムに「明治アーモンドチョコレート 香るカカオ」を認識させている。カメラの映像にこのチョコレートの箱が映っている時にはアプリケーションは反応し、一方で映っていない時には反応していないことがわかる。



図 1:「明治アーモンドチョコレート 香るカカオ」を識別させた例(下段は Amazon の商品画像)

また図 2 では自動販売機に並ぶ「三ツ矢 100%オレンジミックス」を識別させている。多くの商品が並んでいる中で、このジュースが写っている時にのみしっかりと反応していることがわかる。

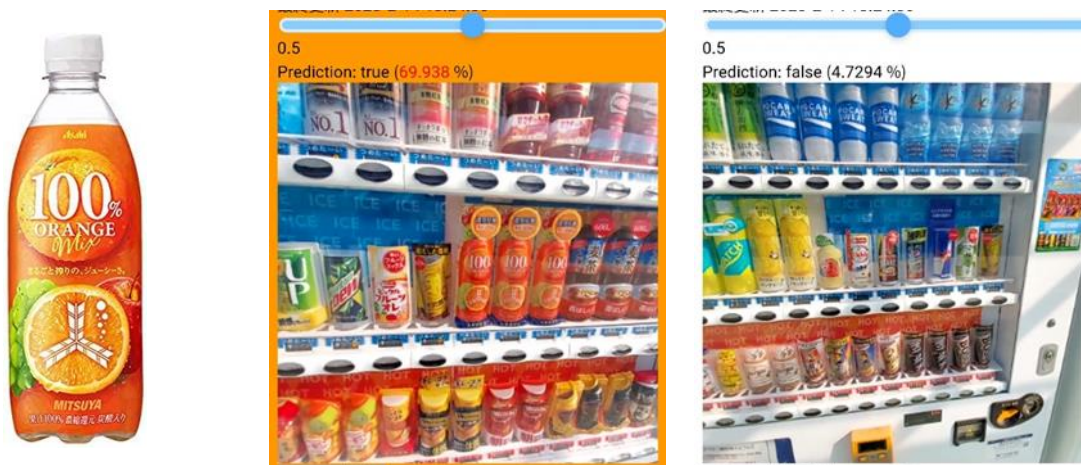


図 2:「三ツ矢 100%オレンジミックス」を識別させた例(左は Amazon の商品画像)

3.2. 商品認識の仕組み

商品の認識のために、図 3 のようなモデルを作成し、使用した。

モデルは商品画像と店内画像(カメラの映像)の 2 つを入力とする。それぞれの画像の特徴量を 2048 次元のベクトルとして取り出し、それらを掛け合わせ、全結合層に通すことで最終的に 0 以上 1 以下の値を得る。この値は 2 つの画像の類似度として解釈でき、1 に近いほど 2 つの画像に同じものが写っている可能性が高いとモデルが判断したといえる。物体の特徴そのものを覚えるのではなく、2 枚の画像の類似度を判断できるようなモデルを用いることで、リアルタイムな物体認識を実現している。

画像を特徴抽出器にかけるのは、モデルは画像情報をそのまま取り扱うことが難しいからである。特徴抽出器にかけ画像内の情報を上手く取り出すことで、モデルが情報を取り扱いやすくしている。したがってこのモデルでは上手く特徴を取り出せるような特徴抽出器を用いることが非常に重要である。

性能の高い特徴抽出器は計算コストが大きく、スマートフォンのようなリソースが限られた環境で用いることは難しい。そこで本プロジェクトではこの特徴抽出器の計算コストを小さくする工夫を行った。特徴量を計算する過程において、従来は計算コストの大きい方法のみを用いていたが、その一部をより計算コストの小さいものに置き換えて結果を再現することで計算コストの削減を図った(GhostNet, 図 4)。その結果、計算量は従来の半分まで削減することができ、スマートフォン上でも充分高速に計算を行えるようになった。

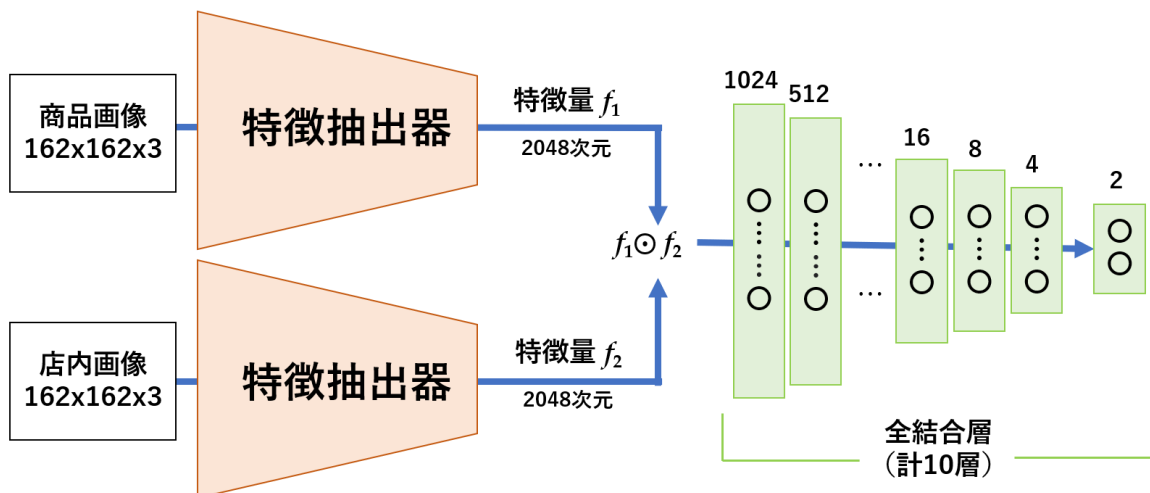


図 3 モデルの概略図

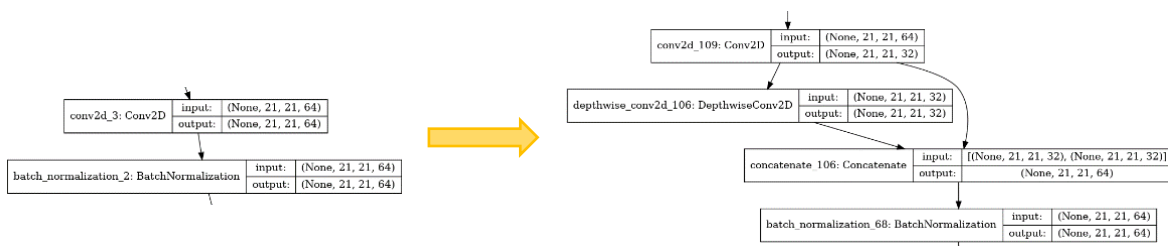


図 4 従来の特徴抽出器の中身の一部(左)と改善した特徴抽出器の中身の一部(右)。全てを計算コストの大きい方法(Conv2D)ではなく、計算コストの小さい方法(DepthwiseConv2D)に一部置き換えることによって、精度を保ちつつ計算コストを抑えることに成功している。

4. 従来の技術(または機能)との相違

本アプリケーションは、「1枚の画像を入力することでその物体を即座に認識できるようになる」という点に最大の特徴がある。

先述のように、ある物体をコンピュータに認識させるために、従来技術ではその物体の特徴をコンピュータに学習してもらう必要があった。図2のように「三ツ矢 100%オレンジミックス」を認識させるためには「三ツ矢 100%オレンジミックス」の画像を大量に用意する必要があり、更にその画像を使ってコンピュータが特徴を学習するためにも時間が必要で、このジュースを飲みたいとユーザが思ったその場で使える技術ではなかった。一方で本アプリケーションでは商品画像は1枚あればよく、この画像を入力したその場で即座にアプリケーションはその物体を汎用的に認識できるようになることが大きな利点である。

またスマートフォンで誰でも気軽に利用できることも大きな相違である。従来技術は基本的に高性能なコンピュータを用いて認識を行うことを前提としており、スマートフォンのようなリソースが限られた環境では基本的に利用できなかった。また従来技術を利用する環境を整えるためにも知識が必要であり、誰でも利用できるものとは言えなかった。本アプリケーションはそのような点を克服し、サイトにアクセスするだけでスマートフォンからでも即座に利用できるようになった。

5. 期待される効果

本アプリケーションを利用することで、多くの人を物探しの煩わしさから解放する効果が期待できる。同じ場所を何度も見直す必要がなくなり、探し物にかかる時間を短縮できるようになる。また商品探しにおいて、依然我々の生活に影を落とすコロナ禍によって、店員さんに商品の場所を尋ねることが憚られたり、そもそも尋ねる店員さんが少なかったりといった状況が続いているが、そのような状況においても商品探しをすることを可能とする。

また見落としを防止することで安全な社会へ一役買う効果も期待できる。つい見逃しがちな一方通行の標識を本アプリケーションが検知するようにすることで逆走を防止したり、お店の看板などを探していると他の部分に対する注意が散漫になってしまいがちだが、本アプリケーションが代わりに探すことで周りへの注意をおろそかにしないようにしたりといったことが可能になる。本アプリケーションの利用によってより安全な社会へ近づくことに繋がる。

6. 普及(または活用)の見通し

今後もより認識精度の高いシステムを目指して開発を続けていく予定である。また単に写っているか否かのみではなくて、画面のどこに写っているかといった情報を提供できるようにしていきたい。更に直感的に操作方法がわかるようなUIにするなど、UXに対する改善も続ける。

上記に加え、物探しに際して本アプリケーションを利用した使い勝手などをまとめ、論文として発表したいとも考えている。先述のように対象の物体を学習した物体認識・物体検出は既に多くの研究が発表されてきたが、本アプリケーションが用いている未学習の物体を認識・検出するタスクは先行研究に乏しい。本アプリケーションの利用によって研究界隈全体を盛り上げ、物探しアプリケーションの普及に繋げていきたい。

7. クリエータ名(所属)

長沢 瑛史(静岡大学 大学院総合科学技術研究科 工学専攻)

(参考)関連 URL

FineTop: <https://fine-top.sakura.ne.jp/>