

EC サイトにおける視覚障がい者向け商品説明のための 画像認識に基づく説明文自動生成

田中 駿哉† 寺澤 卓也†

東京工科大学メディア学部†

1. はじめに

近年のインターネットは最新技術の進歩で性別、年齢、体の不自由問わず誰でも便利に利用できるようになった。視覚障がい者はインターネットを利用する際、画面読み上げソフトを使用していることが多いが、現在展開されている画面読み上げソフトは画像を読み上げないため、画像の情報を理解するのが困難である。

本研究では、視覚障がい者が一番利用しているネットサービスである EC サイトに着目し、衣類の EC サイトを利用する視覚障がい者が商品画像の説明を音声で聞いたその時に購買の判断ができるような視覚障がい者向けの商品紹介文を生成するシステムの検討を目的とする。

2. 関連研究

立野ら[1]は既存の画像説明文生成手法が視覚障がい者支援を目的とした衣服画像の説明文として有効であるか検証を行った。トップスを対象とし、色や柄など商品の特徴を 10 項目でラベル付けを行った後、的確に抽出するためにファインチューニングを実施した。その結果として約 8 割の出力で正しい結果を得ることが出来、画像特徴量と単語がうまく結びついていることを確認した。しかし服を選ぶ際に必要となる装飾や色の情報が不足しているため、音声で聞いたときにその商品をイメージ出来るような文章になっていない。

3. 提案手法

本研究では商品の説明に必要な「商品の種類」、「デザイン」、「モデルの性別」、「服のパーツ」を各ブロックに分け、各々で画像分類、キーワードの決定を行う。以下、商品の種類を示す「服装ブロック」、色を示す「色彩ブロック」、モデルの性別を示す「性別ブロック」、柄やポケットなどを示す「デザインブ

ック」と命名する。全体の流れを図 1 に示す。

ブロック毎に処理を行うメリットは各ブロック専門のタグのみを付与することができ、1 枚の画像に対し付与するタグ数を減らすことが可能なことである。また、色に特化した分類器や性別に特化した分類器を使用することで、専門の処理を行うことが可能となり、マルチラベルで行うよりも高い精度で出力できると考えられる。

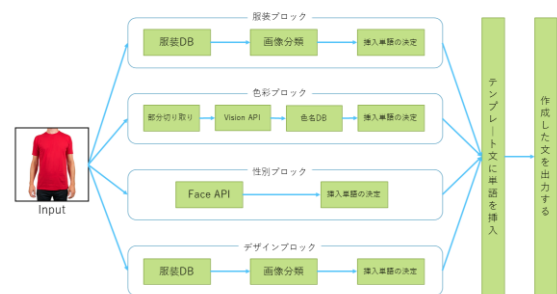


図 1 提案手法の大まかな流れ

4. 実装

色彩ブロックでは Google Cloud Vision API を用いて色彩抽出を行い、RGB を出力する。その後独自に作成した色名データベースにより色名の検索を行い、適した色名を出力する。色名は視覚障がい者がイメージしやすい色である「ピンク系、赤系、橙系、黄系、黄緑系、緑系、青緑系、青系、青紫系、紫系、白系、黒系、灰色」と設定した[2]。

性別ブロックでは MS Face API を用いて性別の判別を行う。性別を判別することで対象商品が男性用か女性用か判断することが出来る。出力は「男性、女性、なし」と設定した。

服装ブロック、デザインブロックでは畳み込みニューラルネットワーク (Convolutional Neural Network: CNN) の ResNet110[3]を用いて画像分類を行った。入力層は RGB の 3 チャンネルで $3 \times 64 \times 64$ に設定し、バッチサイズは 128 で行った。データセットは Liu らが作成した大規模服装データベース Deep Fashion[4]にある服装画像を使用した。Deep Fashion は衣服のカテゴリーやデザインなど 80 万枚を超える多様な衣服画像で構成

A method for creating captions for screen-reader users based on image recognition

†Shunya Tanaka, Takuya Terasawa,

School of Media Science, Tokyo University of Technology

されている。

服装画像には全身が写っているものやトップス、ボトムスが一緒に写っているものもある。そのため服装ブロックでは図2の処理を行う。まず、トップスとボトムスを Google Cloud Vision API で判別を行う。もしボトムスと判断した場合はボトムス用モデルを採用する。もしトップスと判断した場合はさらにインナーかアウターかを再度 Google Cloud Vision API で判別する。もしアウターと判断した場合はアウター用モデルを採用し、インナーと判断した場合はインナー用モデルを採用する。なお、3つのモデルには Deep Fashion の画像から商品種別と商品画像がセットになったデータセットを作成し学習させた。

デザインブロックは留め具、ポケットの有無、袖の有無、襟、柄の5つの分類するデータセットを作成した。

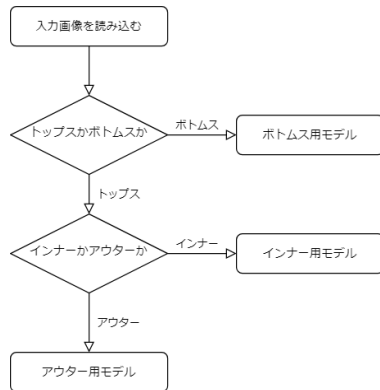


図2 服装ブロックのフローチャート

各ブロックの処理をまとめて実行し、理解できる文章が生成されるかを確認するためテストを実施した。表1に結果を示す。

表1 提案手法の出力結果

入力した商品画像	商品種別	出力した文（誤判定部分に下線）
	タートルネック	これは無地のチノパンです。モデルはなしで、色は黒系です。デザインについて留め具はなし、ポケットはなしです。
	カーディガン	これは無地のカーディガンです。モデルはなしで、色は黄系です。デザインについて留め具はなし、ポケットはなし、袖はなし、襟は丸首です。



ジーパン これは無地のカプリパンツです。モデルはなしで、色はピンク系です。デザインについて留め具はなし、ポケットはなしです。

この結果から商品単体で写っている画像の服装部分は適切に出力できているが、それ以外の服装部分はトップス、ボトムスを分類する時点でボトムスと判定したため、想定していた文章を生成することが出来なかった。また服装部分の分類が失敗している影響で、色彩の認識も適切にできていないところもあった。表1を含め、任意に選択した10枚の画像で確認したが、適切と思われる文章は一つも生成できなかった。

5. まとめ

本研究は視覚障がい者が一番利用している衣類のECサイトに着目し、ECサイトを利用する視覚障がい者が商品画像の説明を音声で聞いて購買の判断ができるような、視覚障がい者向けの商品紹介文を生成するシステムを検討した。今回生成された商品紹介文は服装やデザイン、性別などが期待通りには出力されなかったため、適切な商品紹介文が生成できなかった。

その原因はトップス、ボトムスの分類が上手くできなかったこと、データセットに使用した画像に工夫の余地があったことであると考えられる。今後の課題として、トップス、ボトムスの分類とデータセットの改善を行い、表1の結果よりも適切な文章が出力できるか再度検証実験を行う予定である。

参考文献

- [1] 立野貴梨, 高木昇, 澤井圭, 増田寛之, 本吉達郎. 視覚障害者支援を目的とした衣服画像説明文生成手法. 第36回ファジィシステムシンポジウム講演論文集. 2020, 36th, p.473-476.
- [2] 奥寺沙織, 佐川賢. “全盲視覚障害者の色彩心理構造—全盲視覚障害者に衣服の色を伝えるために—”. 日本女子大学紀要 家政学部. 2014, 第61号, p81-90.
- [3] K. He, X. Zhang, S. Ren and J. Sun, "Deep Residual Learning for Image Recognition," in 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2016, p. 770-778.
- [4] Z. Liu, P. Luo, S. Qiu, X. Wang and X. Tang, "DeepFashion: Powering Robust Clothes Recognition and Retrieval with Rich Annotations," in 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2016, p. 1096-1104.