

## 深層学習技術に適応的データ拡張手法を用いて誤認識リスクを低減した値引きシール認識機能

Discount Label Recognition Function Using Adaptive Data Augmentation Technique to Reduce Risk of Misrecognition in Deep Neural Networks

竹野 唯志 TAKENO Yuishi 信岡 哲也 NOBUOKA Tetsuya

近年、小売業界ではセルフレジが導入され、購入者自身でタブレットやスマートフォンのカメラを使って商品登録する環境が普及し始めた。商品の価格情報には、バーコード以外に店舗によって異なるデザインの値引きシールが貼られたものもあり、値引きシールの認識結果は会計金額に影響するので誤認識リスクの低減が求められる。

東芝テック(株)は、端末上で軽量に動作する、深層学習による値引きシール認識機能の開発を進めている。今回、従来手法に比べて誤認識リスクを更に低減するために、値引きシールのデザインに応じた動的なデータ拡張アルゴリズムを採用し、低コスト化のためにデザインの事前知識に頼らない、独自の適応的データ拡張手法を開発した。これにより、様々なデザインの値引きシールに短納期及び低コストで対応できるようになる。

Self-checkout systems, which provide an environment allowing customers to register purchased products by themselves based on image information captured by a camera on a terminal such as a tablet or smartphone, have become widely disseminated in the retail industry in recent years. However, as price information includes discount labels with different designs of individual stores in addition to the price barcode label of each product, it is essential to reduce the risk of discount label misrecognition so as to avoid mistakes in the amount payable for purchases.

Toshiba Tec Corporation has been developing a discount label recognition function for lightweight applications running on tablets and smartphones. We have now further enhanced the discount label recognition function using an adaptive data augmentation technique instead of the static data augmentation technique conventionally used for training deep neural networks. This technique incorporates the following features: (1) sufficient reduction of misrecognition risk through the application of a dynamic data augmentation algorithm appropriate for discount label designs and (2) adaptive image processing without the use of previous knowledge related to discount label designs. These features make it possible to rapidly accommodate various types of discount label designs at low cost.

### 1. まえがき

近年、ショッピングカートにタブレットが搭載されたカートPOS (Point of Sales) や、来店客自身のスマートフォンを利用するスマートフォンPOSが普及し始めている<sup>(1)</sup>。カートPOSの場合は、店舗側で準備したタブレットがショッピングカートに固定されており、通常はチェックアウトゾーンで行う商品登録作業を、売り場で購入商品を選定しながら来店客が自ら端末を使用して完了させる。商品登録作業は、購入したい商品のバーコードを端末のカメラに読み取らせ、POSアプリケーションのリストに商品を追加する作業である。スマートフォンPOSの場合は、来店客自身が所有するスマートフォンに、事前に専用アプリケーションをインストールしておくことで、カートPOSと同様に商品登録作業を実現する。システムによっては、決済まで端末上で完了できる。

カートPOSやスマートフォンPOSは、店舗側にとっては、売り場で来店客個別にきめ細かなプロモーションを実施できること、及びチェックアウトゾーンに確保するスペースや配置

する店員数を削減できるというメリットがある。来店客には、自分に合ったレコメンドを受けられることや、レジ待ちをする必要がないというメリットがある。セルフレジが普及した現在、来店客自身に商品登録作業を課すことは受け入れられやすくなっており、今後、カートPOSやスマートフォンPOSの普及が加速していくことが予想される。

東芝テック(株)は、値引きシールの貼り付けられた商品をカートPOSやスマートフォンPOSでも扱えるようにするために、端末上で軽量に動作する、深層学習技術を活用した値引きシール認識機能を開発し、製品への搭載を予定している。端末に搭載されたカメラで撮影した画像の中から認識対象となる値引きシールを探し出し、10円引きや30%引きといった割引額を認識する処理を端末上で実行できる。スマートフォンPOSの場合、来店客自身のスマートフォンが使用されるため、端末の処理性能やカメラ性能にばらつきが生じる。様々な端末をサポートできるように、認識処理の軽量化と、撮影環境へのロバスト性を確保した。

深層学習技術を活用する場合、深層学習モデルに学習さ

せるための大量のデータが必要であり、開発コスト増の要因になる。一方で、小売業者は店舗独自の値引きシールの認識を希望することが多く、認識精度を高めるためには、更に多くの学習データが欠かせない。このニーズに低価格で対応するためには、学習データ準備の低コスト化は必須の課題である。

ここでは、撮影環境へのロバスト性確保・学習データ準備の低コスト化・誤認識リスク低減のために実施している静的なデータ拡張手法と、今回新たに開発した、認識対象となる値引きシールのデザインの事前知識に頼らない適応的データ拡張手法について述べる。

## 2. 分類問題へのアプローチ

小売業界では、賞味期限が近づいてきた商品に値引きシールを貼り、来店客の購買意欲を喚起して消費を促すことが広く行われている。商品によって割引額が異なることや、同じ商品でも賞味期限までの残り時間によって割引額が異なることがある。したがって、値引きシール認識機能には、商品に値引きシールが貼り付けられていることの検知に加えて、割引額の特定も求められる。値引きシール認識の結果は、会計金額に影響するため、誤認識は避けなければならない。

来店客が端末に商品をかざしている間、端末のカメラが撮影した画像データを次々と取得して認識させることができるので、誤認識するリスクのあるときはリジェクトし、正解できる可能性が高いときだけ認識結果を出力する。リジェクトの頻度が高いと読み取りの反応性や即時性が低下するが、無理に認識結果を出力して誤認識するリスクを負うのではなく、リジェクトすることを優先する方がよい。

当社の値引きシール認識機能の処理フローを、図1に示す。前段では撮影画像の中から値引きシールの領域を特定し、後段では分類問題のアプローチで割引額を特定する。つまり、あらかじめ10円引きや30%引きといった割引額それぞれに対応するクラス(カテゴリー)を定義しておき、値引きシールがどのクラスに分類されるのかを、深層学習技術を

活用して算出する。ただし、上述したように適切にリジェクトすることが求められるため、割引額に対応するクラスに加えて、リジェクトに対応するリジェクトクラスを追加定義しておく。認識処理によってリジェクトクラスに分類された場合は、次の撮影画像データでの認識を試みる。

## 3. データ拡張手法

この章では、当社がこれまで実施してきた静的なデータ拡張手法について述べる。

### 3.1 学習データの取得

実際に店舗で使用される端末は多種多様であることが想定されるが、全ての端末を準備することは不可能であるため、複数の候補機種を選定して学習用画像を撮影する。各端末を使用して撮影した値引きシールの画像を、図2に示す。撮影時に値引きシールの割引額を記録しているため、取得した全ての画像には正解ラベルが付与された状態である。また、背景を黒地としているため、値引きシール領域だけを容易に抽出できる。

### 3.2 CGを使った加工・合成処理によるデータ拡張

各端末を使用して撮影した値引きシールの画像に対して、値引きシールの一部分を隠した状態を模擬する遮蔽(しゃへい)加工や、店内の照明の影響を模擬する照明加工といった各種加工処理を、CG(Computer Graphics)の技術を駆使して施す。現実には起こり得るバリエーションを網羅した様々な加工処理を施して拡張した学習データを利用することで、撮影環境へのロバスト性向上や学習データ準備の低コスト化を実現する。様々な加工処理を施した後、あらかじめ用意してある背景画像上に合成する。加工前の撮影画像と、各種加工後の画像、及び合成後の画像を背景画像上に合成した画像の例を、図3に示す。

リジェクトクラスを正解とするリジェクト用サンプルとしてどのような画像を選定するかは、誤認識リスク低減に寄与する重要な要素である。リジェクト用サンプルは、割引額の特定を困難にするために、割引額の特定に必要な情報が写っているエリアに対して破壊的加工処理を施して生成する。破



図1. 値引きシール認識機能の処理フロー

端末で撮影した画像のシール領域を特定し、割引額のカテゴリを特定する。

Flow of processes to recognize discount label



図2. スマートフォンで撮影した値引きシールの画像例

複数種類の端末で撮影した画像に、割引額の正解ラベルを付与して、学習データとする。

Examples of images of discount labels captured by smartphone

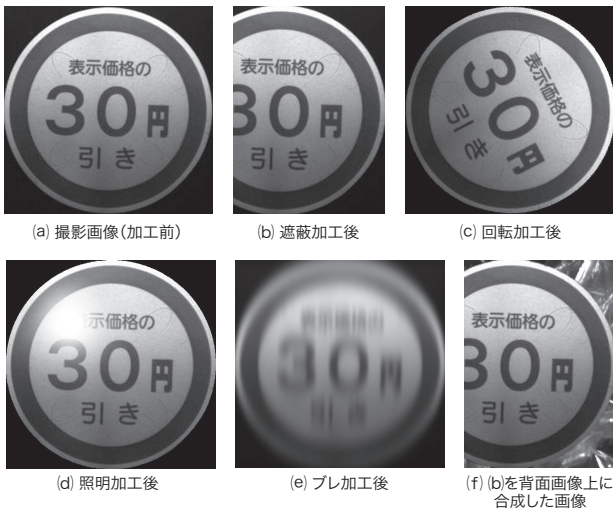


図3. 撮影画像の加工・合成処理の例

従来の静的なデータ拡張手法では、撮影画像に対して、各種加工処理を適用し、背景画像上に合成して、データを生成する。

Examples of processed and synthesized images of discount label

壊的加工処理は、例えば図3(b)の遮蔽加工処理のように、認識に必要な情報を欠損させるような加工処理を指す。また、図3(d)の照明加工処理の例では、強い照明で白飛びして元のテキスト情報が失われる場合は壊的加工処理に該当するが、若干の陰影を与える程度の場合は壊的加工処理には該当しない。例えば、図3(d)で示される「表」の文字周辺だけが白飛びしたサンプルは、割引額特定に必要な情報が欠損していないため、リジェクト用サンプルではなく30円引きクラスのサンプルとして扱う。

今後、店舗ごとに異なる値引きシールに順次対応していくことを想定すると、値引きシール画像の中のどこに割引額の特定に必要な情報が写っているか、というデザイン(レイアウト)に関する事前知識がなくても、リジェクト用サンプルを生成できるようにする必要がある。例えば、遮蔽加工処理において、どのエリアを遮蔽してリジェクト用サンプルにするべきかを、事前知識なしにデザインに合わせて動的に決める必要がある。

#### 4. 適応的データ拡張手法

値引きシールのデザインに関する事前知識がなくても、加工処理後のサンプルをリジェクト用サンプルとして扱うかどうかを適切に判断するために、適応的データ拡張手法を開発した。

データ拡張で生成したサンプルをリジェクト用サンプルにするべきかどうかは、壊的加工処理によって認識に必要な情報が欠損しているかどうかで判断できる。例えば、10

円引きの画像を10%引きや15円引きと区別するためには、「1」、「0」、及び「円」という文字が印刷されたエリアが重要であり、ほかのエリアは重要ではない(注1)。値引きシールのデザインの事前知識がなくても重要なエリアを特定するために、CNN(Convolutional Neural Network)が注目したエリアをヒートマップで可視化する技術を活用する。この可視化技術として、Grad-CAM<sup>(2)</sup>を適用した。

適応的データ拡張手法の概要は、以下のとおりである。

- (1) 端末で撮影した画像を、データ拡張しないでそのまま使用してCNNに学習させる。この段階では、リジェクト用サンプルは生成しない。
- (2) 端末で撮影した各画像に対して、学習したCNNが注目するエリアを表すヒートマップを、Grad-CAMで生成する。
- (3) Grad-CAMで生成したヒートマップに対して、ノイズ除去フィルターや二値化などの処理を適用し、注目エリアを表す注目マップを生成する。
- (4) データ拡張を実行し、壊的加工処理によって注目マップが示すエリア内の情報が欠損した場合に、正解ラベルをリジェクトクラスに変更する。

図4に、適応的データ拡張手法の具体例を示す。

まず、3.1節で述べたように、複数のスマートフォンで認識対象となる値引きシールを撮影する(図4(a))。各画像には割引額に対応するクラスの正解ラベルが設定されている。この画像と正解ラベルを学習データとしてCNNに学習させる。この際に加工・合成処理といったデータ拡張は行わず、生の撮影画像をそのまま使用する。これは、本番の学習の前に一度だけ実行する事前準備の学習である。

学習完了後、学習に使用した生の撮影画像それぞれに対して、学習したCNNが注目するエリアを表すヒートマップをGrad-CAMを適用して生成する。生成したヒートマップはそのままでは扱いにくいので、ノイズ除去フィルターや二値化などの処理を適用して整え、各撮影画像の中のどこが注目エリアなのかを表す注目マップを生成する(図4(b))。図4(c)は、撮影画像上に注目マップを重畳して位置関係を分かりやすくした図である。撮影画像に、3.2節で述べた加工・合成処理を加える。その際に、割引額特定に必要な情報が印刷されたエリア内に壊的加工処理が行われたかどうかを、注目マップを用いて判定する。加工処理が行われた範囲が注目マップのエリア外だけの場合は、正解ラベルを変更せずに80円引きクラスのサンプルとして扱う(図4(d))。加工処理が行われた範囲が注目マップのエリア内に及ぶ場合

(注1) 真贋(しんがん)判定を目的とする場合は、この限りではない。

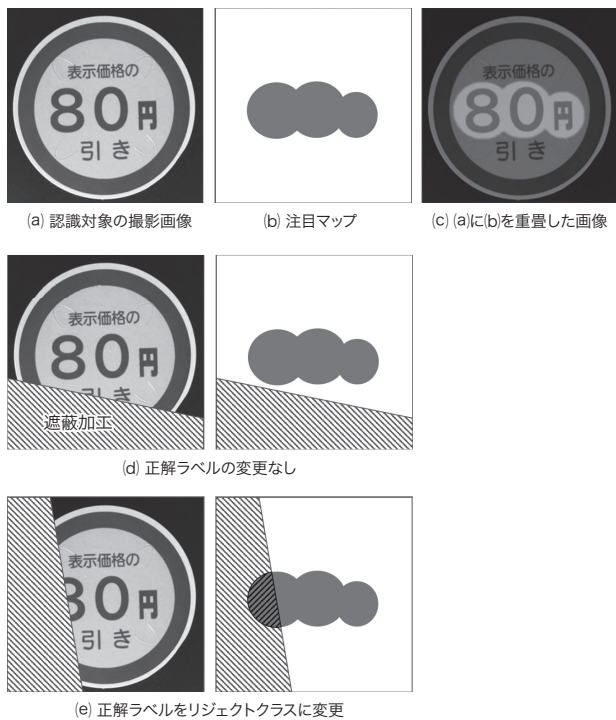


図4. 適応的データ拡張手法の具体例

注目マップのエリアと加工処理した部分の重なり具合から破壊的加工かどうかを判定し、適切なクラスの正解ラベルを付与する。

Examples of application of adaptive data augmentation method

は、破壊的加工処理と判断して正解ラベルをリジェクトクラスに変更し、リジェクト用サンプルとして扱う(図4(e))。

このように、適応的データ拡張手法で作成した学習データを使用して、実使用時の学習を実行する。

## 5. 効果検証

適応的データ拡張手法の効果を、静的なデータ拡張手法を比較対象として、検証した。

ここでは、円形の値引きシール(直径:  $R$ )に、加工処理として図3(b)の遮蔽加工だけを実施し、遮蔽された部分と遮蔽されない部分の境界が直線である場合だけを扱う。遮蔽の度合いを表すパラメータとして、遮蔽率を導入した。遮蔽率は、遮蔽された部分と遮蔽されない部分にある境界線の垂直二等分線上における、遮蔽された部分の輪郭線から境界線までの長さを遮蔽量  $a$  として、 $a/R$  と定義した(図5)。遮蔽の方向はランダムに設定し、今回の検証パラメータからは除外した。

比較対象とする静的データ拡張手法では、遮蔽率がしきい値以下の場合には実際の割引額に対応するクラスの学習データとして扱い、遮蔽率がしきい値を超える場合はリジェクト用サンプルとして扱うようにした。しきい値は、0から0.5

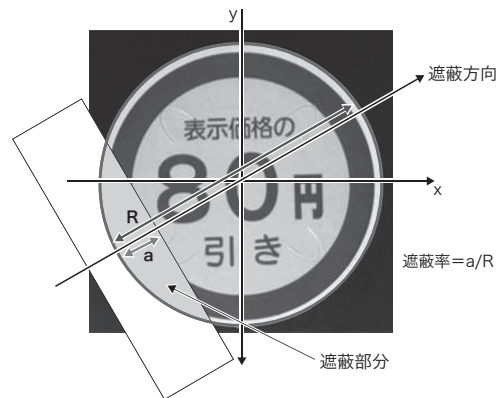


図5. 遮蔽率の定義

遮蔽の度合いを表すパラメータとして、遮蔽率を定義した。

Definition of cover rate

までを8分割して9通り設定した。例えば、しきい値0では、遮蔽加工をしない場合だけを実際の割引額に対応するクラスの学習データとして扱い、わずかでも遮蔽加工をしたもの(遮蔽率  $> 0$ )は全てリジェクト用サンプルとして扱う、ということである。9通りのしきい値に対して、それぞれモデルの学習を実行した。今回の検証では、生成された画像を目視で確認してリジェクト用サンプルとするかどうかを判断する作業は省略し、しきい値を超えたかどうかによって自動的に振り分けた。

しきい値ごとに静的データ拡張手法を用いて作成した学習データによる9通りのモデルと、4章の適応的データ拡張手法を用いて作成した学習データによるモデルを合わせた、計10通りのモデルについて、認識精度を評価した。

評価用データセットは、学習時と同じ条件の遮蔽加工だけを施して、9通りの遮蔽率のデータが、均等の頻度で発生するように、全部で3,384の評価用データを生成した。評価用データセットでは、正解データとしてリジェクトクラスは設定せず、割引額が判別不能な画像であっても実際の割引額に対応するクラスを正解とした。分類結果が正解データと一致すれば正解、分類結果がリジェクトクラスであればリジェクト、それ以外であれば不正解として集計し、全評価用画像数3,384で割って、正解率、リジェクト率、及び不正解率を算出した。

図6に、静的データ拡張手法のしきい値ごとの9通りのモデルと、適応的データ拡張手法のモデルを用いて、評価用データセットを分類したときの、正解率と不正解率を示す。

静的データ拡張手法では、しきい値を大きくすると、正解率が大幅に増加する。これはリジェクト率が低減するからである。一方、不正解率もしきい値の増加とともに増加する傾

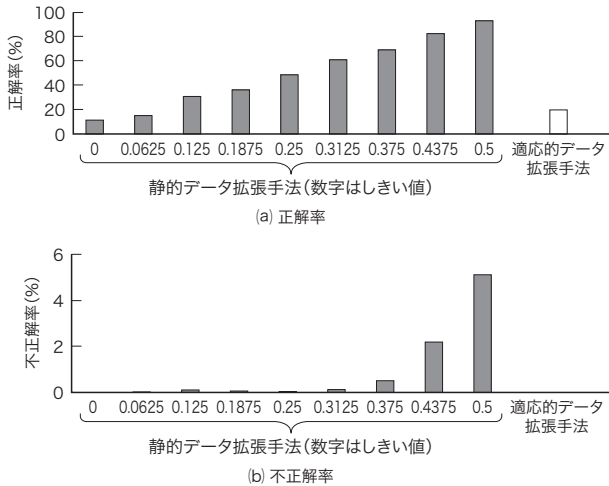


図6. 適応的データ拡張手法と静的データ拡張手法の効果検証結果

従来の静的データ拡張手法は、しきい値によって正解率・不正解率が変化するが、適応的データ拡張手法は、しきい値の設定をしなくても、不正解率を0に抑えながら高い正解率を示した。

Results of comparative evaluation of conventional and adaptive data augmentation methods

向が見られた。また、しきい値を0.0625と0.125の間どこかに設定すれば不正解率を0に抑えながら、正解率を最大化できると推定される。

一方、適応的データ拡張手法の不正解率は0であり、静的データ拡張手法の不正解率0の場合に比べて、高い正解率を達成している。しきい値を値引きシールのデザインに合わせて最適化しなくても、不正解率を抑えて高い正解率を達成できることが検証できた。

実際には、加工処理は様々なパターンがあり、値引きシールのデザインに合わせて都度全てのパラメーターを最適化することは現実的ではない。このような中で、適応的データ拡張手法が効果を発揮する。

## 6. あとがき

従来の静的なデータ拡張手法の課題を解決する適応的データ拡張手法について述べ、その効果の一部を検証した。適応的データ拡張手法を用いることで、認識対象とする新しい値引きシールのデザインの事前知識がなくても、学習データを自動生成できる。

人手に頼っていた調整項目を自動化し、店舗ごとに異なる値引きシールにも短納期で対応できる商品を、提供していく。

## 文献

- (1) 東芝テック(株), "ELERA 売場移動型セルフレジシステム ピピットセルフ". <<https://www.toshibatec.co.jp/products/pos/pipit/>>, (参照 2022-04-25).
- (2) Selvaraju R. et al. "Grad-CAM: Visual Explanations from Deep Networks via Gradient-Based Localization". IEEE International Conference on Computer Vision (ICCV), Venice, 2017-10, IEEE. 10.1109/ICCV.2017.74.



竹野 唯志 TAKENO Yuishi, Ph.D.  
東芝テック(株)  
技術戦略部 リサーチ&開発センター  
博士(工学)  
Toshiba Tec Corp.



信岡 哲也 NOBUOKA Tetsuya  
東芝テック(株)  
リテール・ソリューション事業本部 デバイス企画開発センター  
Toshiba Tec Corp.