5N-01

CNN 特徴量の列要素 L2 制約とバイナリ化ベクトルを用いた画像検索の性能向上

鹿島崇† 金子直史† 鷲見和彦†
青山学院大学理工学部情報テクノロジー学科†

1. はじめに

近年，漫画などの著作物を Web 上へ無断で掲載し，権利元である作者や出版社らに不利益をもたらす違法アップロードが蔓延している．既に大規模な違法掲載サイトがいくつか摘発されているが，それに代わる新たな Web サイトが現在も増え続けているのが現状である．こうした違法コンテンツを早期に摘発するためには自動検知技術が必要であると考えた．

アップロードされる漫画は電子書籍や，発売前のネタバレを目的とした漫画の撮影画像が数多くを占めている．電子書籍や高品質にスキャンした画像などであれば，既に存在する電子データとの差異がほとんどないことから違法コンテンツの検知は容易であると考えられる．しかし，撮影画像は外乱などの影響によって画像特徴量が，電子書籍から得られる特徴量と大きく異なる場合がある．その結果，同一コンテンツであるにも関わらず別のコンテンツとして識別されてしまう問題が生じる．

そこで本研究では，畳み込みニューラルネットワーク(CNN)の特徴量に基づいた画像検索による，電子書籍と撮影画像間の同一コンテンツ認識を行った．CNN モデルに提案手法を用いた学習を行うことで ImageNet で pre-training したモデルや，距離学習，Batch Normalization などの手法を上回る検索精度を達成した．

2. 提案手法

バッチ方向に L2 正規化を行う，L2-Batch Normalization Layer を提案する．

2.1 L2-Batch Normalization Layer

CNN モデルから得られる画像 1 枚の n 次元特徴ベクトル(ReLU 関数の出力とする)を式(1)に示す．

$$x_{vector} = (x_1 \quad x_2 \quad \cdots \quad x_n) \qquad \cdots (1)$$

学習時はバッチを処理するため，バッチサイズを m としたとき，式(2)のような $m \times n$ 行列が定義される．

$$x_{batch} = \begin{pmatrix} x_{11} & x_{12} & \cdots & x_{1n} \\ \vdots & & \ddots & \vdots \\ x_{m1} & x_{m2} & \cdots & x_{mn} \end{pmatrix} \qquad \cdots (2)$$

ここで，式(2)のバッチ方向に L2 正規化を行う．式(2)は二次元テンソルであるためバッチは列に該当し，各列の単位ベクトル化によって要素値は範囲$[0,\ 1]$に制約される．

バッチ方向の L2 正規化は以下のように定義される．

$$\varepsilon = 1e-7,\ \ max(a,\ \varepsilon) = \begin{cases} \varepsilon, & a \le \varepsilon \\ a, & a > \varepsilon \end{cases} \text{ とする．}$$

学習時：

$$= \begin{bmatrix} \frac{x_{11}}{max\left(\sqrt[2]{\sum_{i=1}^{m} x_{i1}{}^2},\ \varepsilon\right)} & \cdots & \frac{x_{1n}}{max\left(\sqrt[2]{\sum_{i=1}^{m} x_{in}{}^2},\ \varepsilon\right)} \\ \vdots & \ddots & \vdots \\ \frac{x_{m1}}{max\left(\sqrt[2]{\sum_{i=1}^{m} x_{i1}{}^2},\ \varepsilon\right)} & \cdots & \frac{x_{mn}}{max\left(\sqrt[2]{\sum_{i=1}^{m} x_{in}{}^2},\ \varepsilon\right)} \end{bmatrix} \cdots (3)$$

次に，入力画像 1 枚に対して推論を行う場合の処理について説明する．式(1)のベクトルは ReLU 関数の出力であるため，L2-Batch Normalization Layer から特徴抽出を行うことで式(1)の特徴ベクトルを二値化したバイナリベクトルが得られる．

推論時：

$$= \left[\frac{x_1}{\max\left(\sqrt[2]{x_1{}^2},\ \varepsilon\right)} \quad \frac{x_2}{\max\left(\sqrt[2]{x_2{}^2},\ \varepsilon\right)} \quad \cdots \quad \frac{x_n}{\max\left(\sqrt[2]{x_n{}^2},\ \varepsilon\right)} \right] \cdots (4)$$

$$\frac{x_t}{\max\left(\sqrt[2]{x_t{}^2},\ \varepsilon\right)} = \begin{cases} 1, & x_t > 0 \\ 0, & x_t = 0 \end{cases} \quad (1 \le t \le n)$$

2.2 提案手法を適用した学習モデル

VGG16[1]をベースに，提案手法を追加したモデル構造を図 1 に示す．VGG16 の畳み込み層は ImageNet による訓練後の重みを利用し，block1 から block4 までの重みを凍結する．学習では block5 と全結合層 fc1，fc2，prediction の fine-tuning を行う．検索に用いる層は類似画像検索の先行研究において fc1 層が検索に適すと報

Improving Image Retrieval Performance Using CNN Feature's L2-constrained Column Elements and Binary Vectors
†Takashi Kashima †Naoshi Kaneko †Kazuhiko Sumi
† Aoyama Gakuin University

告されており[2]，その次の層に L2-Batch Normalization Layer を追加する．

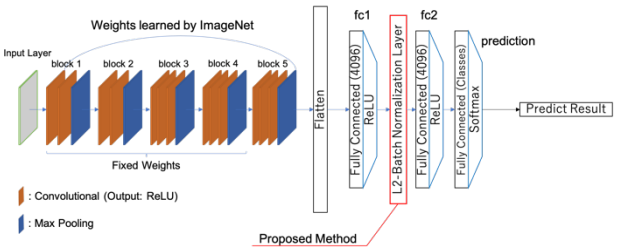

図 1.　提案手法を追加した VGG16

3.　学習

電子書籍と撮影画像の両方が含まれたデータセットで分類学習を行う．学習では，漫画の 1 ページを電子書籍 1，2 枚と複数枚の撮影画像で構成される 1 つのクラスとして扱う．

3.1 学習用データセット

データセットの概要を図 2 に示す．電子書籍は，iPhone 公式アプリ「Books」の無料ダウンロード可能な漫画をスクリーンショットで収集を行った．撮影画像は，漫画の単行本を様々な距離・角度を変更しながら 17 枚から 281 枚の範囲で撮影して収集を行った．画像を(224 × 224)にリサイズし，ラベルを付与した計 75052 枚をデータセットとした．

図 2.　データセットの概要

3.2 訓練

提案手法のほか，提案手法を用いない場合や，距離学習の CosFace，ArcFace，SphereFace，AdaCos，L2-constrained Softmax Loss，特徴抽出層のみを Batch Normalization で正規化した VGG16 をそれぞれ fine-tuning する．75052 枚のうち訓練データ 60250 枚，　撮影画像のみの検証データ 6671 枚，テストデータ 8131 枚，電子書籍のみのテストデータ 601 枚(1 クラスのみ言語別で 2 枚)とする．全モデルの学習はバッチサイズ 20 とし，　30000 イテレーションまで行う．

4.　検索精度評価

ImageNet による pre-train モデルと，提案手法等の学習モデルによる検索精度の測定と比較を行う．分類学習時の 600 クラスに該当しない 522 枚(全て異なるページ)の電子書籍と撮影画像のペア計 1044 枚を用意し，検索精度を評価した．

Manga109[3][4]のページ 21238 枚と，電子書籍か撮影画像の一方の 522 枚を合わせた計 21760 枚を特徴ベクトル化したデータベースを作成し，ペアのもう一方である 522 枚をクエリとする．特徴量は提案手法の場合，式(4)のバイナリベクトル，それ以外は fc1 層の出力を単位ベクトル化したものを使用し，L2 距離で検索を行う．

全数探索による Top-1，Top-5 検索精度を測定した結果を図 3 に示す．提案手法が他の手法を上回り，最高精度を達成した．

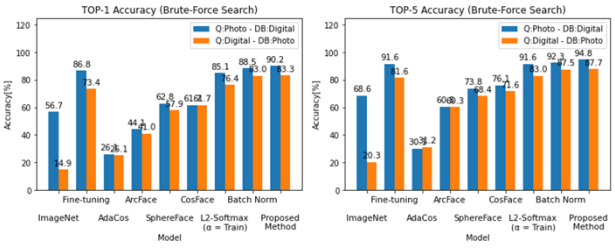

図 3.　Top-1，Top-5 検索精度比較

5.　まとめ

本研究では L2-Batch Normalization Layer を用いた学習と，バイナリベクトルの利用による同一コンテンツ画像検索の性能向上を示した．

実験で得られた検索精度は学習時のクラスに属さない画像を用いた結果であるため，モデルを再学習させることなく，新しく創作される漫画ページに対しても同様のコンテンツ検索が期待できる．

参考文献

[1] K. Simonyan and A. Zisserman, “VERY DEEP CONVOLUTIONAL NETWORKS FOR LARGE-SCALE IMAGE RECOGNITION”, In CVPR, 2015.

[2] A. Babenko et al., “Neural Codes for Image Retrieval”, In ECCV, 2014.

[3] Y. Matsui et al., “Sketch-based Manga Retrieval Using Manga109 Dataset”, Multimedia Tools and Applications, vol. 76, no. 20, pp. 21811–21838, 2017.

[4] K. Aizawa et al., “Building a Manga Dataset “Manga109” with Annotations for Multimedia Applications”, IEEE MultiMedia, vol. 27, no. 2, pp. 8–18, 2020.