

ファッション画像に特化した深層学習によるクラスタリング

小路光, 熊野創一郎, 山崎俊彦

東京大学

1. はじめに

近年、Deep Neural Networks (DNN) の発展によって以前は難しいと考えられてきた複雑なタスクが数多く実現している。また、近年では DNN を用いてファッションに関する画像推薦、クラスタリング、流行予測のようなタスクの研究が盛んに行われている。[5, 3]

特に、ファッション画像のクラスタリングに関する手法は、スタイルに基づいた分類、衣服に関するメタ情報を利用した分類、ファッション画像のみを利用した分類のように様々な手法が提案されている。しかし、画像のみを用いた分類の一般的な手法でファッション画像をクラスタリングしてもうまくいかないことを実験的に確認した。そこで、本研究では人間の感性に基づいたファッション画像のクラスタリングを実現できる DNN の提案を行う。そのモデルと既存モデルを用いて画像の類似検索を行い、結果を比較した。

2. 関連研究

2.1 クラスタリング

データセットをクラスタリングする基本的な手法として K-Means [8]、Mean shift [1] など数多く存在する。近年、ファッションに関するデータセットにおいて、コーディネートに似ているグループに分ける研究が数多く行われている。ファッションアイテムの持つ複数の属性の組み合わせによる「スタイル」を定義し、事前にいくつかのスタイルを設定した上で、教師ありアプローチを用いて各スタイルごとに画像进行分类する研究 [4] や、[9] のようにスタイルを事前に決定せずに自然言語処理分野のトピックモデルによる教師なしクラスタリングによって人間が自然と認識しているファッションスタイルに近い結果を得ることができた研究がある [7]。特に後者の手法は、教師ありクラスタリングによる分類とは違い、正解ラベルを用意することが必要なく、ファッションというトップダウン的に服の種類分けが決まらない分野において有用であると考えられる。

2.2 自己教師あり学習

ラベルの付いたデータセットを作成するのはコストがかかるが、ラベルのついていないデータは常に生成されている。自己教師あり学習は、大量のラベルなしデータを活用する試みである。この手法は、データの潜在的な特徴を DNN

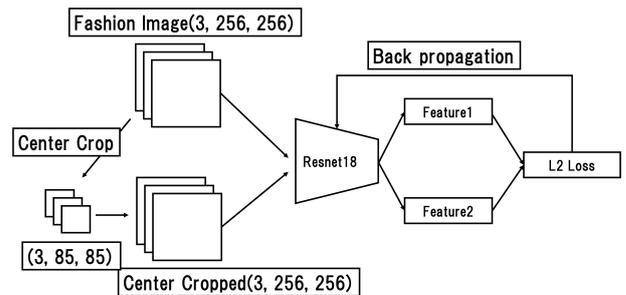


図 1: 提案モデル

に学習させる表現学習で広く用いられている。自己教師ありアプローチはコンピュータビジョン、ロボット制御のような分野で用いられている。[2]

3. 提案手法

本研究では、図 1 のような対照学習モデルを構築した。衣服の質感や柄の特徴量をよりの確に抽出することで今までより性能の良いクラスタリングを実現することを目的とした。そのため、ファッション画像そのものとその画像の中心部を切り出した画像について対照学習を行う。その学習過程を以下の 3 段階に分けて説明する。まず、入力については、未加工のファッション画像とその画像を 3*3 に分割したときの中心部の画像の 2 つを使用する。ここでは画像の中心に衣服が存在しているという仮定をしている。次に、学習は Resnet18 の最終層を取り外したものを使用している。このモデルに 2 枚の画像を入力し、それぞれの画像に対する深層特徴量を抽出する。そして得られた二つの特徴量の平均二乗誤差を計算し、この値を最小とするようにモデルのパラメータを更新する。

4. 実験

クラスタリングに先立って、ファッション画像データセットに対して類似度検索を行った。類似度検索は、ImageNet データセットで事前学習された VGG16 をベースとしたモデル [10] と、本研究で提案するモデルのそれぞれに対して行った。今回の類似検索は図 2 の画像を基準とした。本研究では DeepFashion2 データセットを用いた。DeepFashion2 [6] には衣服に関するデータが収録されている。今回の実験では

trainに50,000枚、validationに30,000枚、testに10,000枚のファッション画像を使用した。



図 2: 類似検索基準画像

4.1 実験結果

提案したモデルの学習は100epochとした。学習が終了したモデルと[10]のモデルそれぞれを用いてDeepFashion2のtestデータ10,000枚に対して深層特徴量を抽出した。データセット内からランダムに選択したファッション画像データに対して類似度検索を行った結果を図3に示す。これによると、我々が提案したモデルの方が直感的には類似画像を抽出できていると伺える。さらに、クラスタリングを行った結果を図に示す。クラスタリングの結果を見ると、今回の基準画像に対しては大きな差異は見られなかった。背景の情報をまだ受けていると考えられる。

5. まとめ

本研究では、衣服のテクスチャに注目した学習を行うことで既存手法に比べて人間の感性に合うクラスタリングを実現するための手法を提案した。また、本実験ではクラスタリングに先立ち、提案手法により、類似検索の結果を既存手法と比較した。さらなる精度向上のために、モデルを改良し質の高いクラスタリングにつなげていきたい。今後は、適切な拡張を選択したり、マルチタスクモデルを構築して実験を行う予定である。

参考文献

- [1] Yizong Cheng. Mean shift, mode seeking, and clustering. *IEEE*, Vol. 17, pp. 790–799, 1995.
- [2] Deepak Pathak et al. Curiosity-driven exploration by self-supervised prediction. In Doina Precup and Yee Whye Teh, editors, *Proceedings of the 34th International Conference on Machine Learning*,



(a) 提案手法 1-5 位



(b) VGG16 ImageNet 事前学習モデル 1-5 位

図 3: 類似度検索の結果

ing, Vol. 70 of *Proceedings of Machine Learning Research*, pp. 2778–2787. PMLR, 06–11 Aug 2017.

- [3] Huizhong Chen et al. Describing clothing by semantic attributes. In *ECCV*, 2012.
- [4] Mohammad Hadi Kiapour et al. Hipster wars: Discovering elements of fashion styles. In *ECCV*, 2014.
- [5] Shuyun Ren et al. Fashion sales forecasting with a panel data-based particle-filter model. *IEEE Transactions on Systems, Man, and Cybernetics: systems*, Vol. 45, pp. 411–421, 2015.
- [6] Yuying Ge et al. Deepfashion2: A versatile benchmark for detection, pose estimation, segmentation and re-identification of clothing images. *CVPR*, pp. 5332–5340, 2019.
- [7] Ziad Al-Halah et al. Modeling fashion influence from photos. *ArXiv*, Vol. abs/2011.09663, , 2020.
- [8] J. A. Hartigan and M. Anthony. Wong. A k-means clustering algorithm. 1979.
- [9] Wei-Lin Hsiao and Kristen Grauman. Learning the latent “look”: Unsupervised discovery of a style-coherent embedding from fashion images. *ICCV*, pp. 4213–4222, 2017.
- [10] Y. Zhang and T. Yamasaki. Style-aware image recommendation for social media marketing. *ACMMM*, 2021.