Computer Vision meets Fashion

山口 光太
Yamaguchi, Kota

AI Lab
Research Scientist
yamaguchi_kota@cyberagent.co.jp

keywords: コンピュータビジョン, 画像認識, ファッション，Street-to-shop，トレンド分析，スタイル生成

Summary

本稿ではコンピュータビジョンをファッションに関連する問題に適用したこれまでの研究を俯瞰する．衣服や属性の画像認識に始まる基本的な問題から，商品検索やスタイルの推薦，ファッションを構成する意味要素の理解，Web データからの学習についての研究事例を，著者がこれまでに取り組んできた事例を含めて解説し，コンピュータビジョンがファッション分野にもたらす研究のこれからについて述べる．

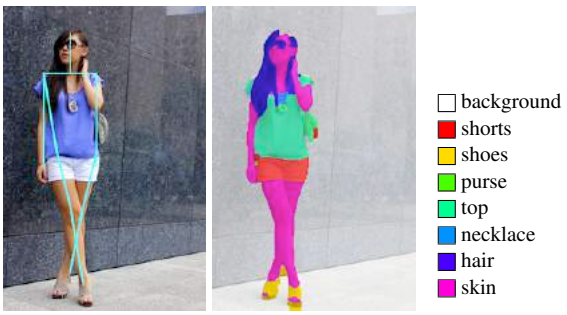
1. は じ め に

機械学習の応用が拡大するにつれ，ファッション分野でもコンピュータビジョンの適用が広がりを見せている．Amazon は IoT デバイスによるパーソナルスタイリストサービスを発表し[*1]，また EC 活用などを目指し数多くのスタートアップ企業（Fashwell，Wide Eyes，VASILY）がファッション画像の認識をコアにしたデータ分析や検索サービス開発を進めている．ファッション業界のマーケットは全世界で 3 兆ドル規模にも達し[*2]，コンピュータビジョンや機械学習がターゲットとするにはまたとない産業分野でもある．日本国内でも機械学習コンペが開催されるなど[*3]，学術研究の面からも近年はファッションを題材にしたものが数多く見られるようになり，国際会議 KDD や ICCV でもファッションを題材にしたワークショップが開催され[*4] [*5]，研究分野の進展も著しい．本稿ではファッションを題材としたコンピュータビジョンのこれまでの研究を俯瞰し，今後の展望について述べる．

2. これまでの研究動向

2·1 衣服の画像認識

コンピュータビジョンが取り扱う最も基本的なファッション認識の問題が衣服の種類の認識（Clothing recognition）である．これは与えられた画像についてどのような衣服であるのかを自動で判別するもので，基本的な衣服のカテゴリ（トップス，パンツ，スカート）の他に V ネック，フレアスカートといった細部のパーツやスタイルの分類は属性認識（Attribute recognition）として取り扱われている [Chen 12, Chen 15, Liu 16a].



(a) 姿勢推定　(b) 領域分割

図 1: 衣服の認識は身体の姿勢と密接に関わる．著者らの手法 [Yamaguchi 12] では姿勢推定結果 (a) を条件入力としてセグメンテーション (b) を求めている．

衣服の画像認識は一般的な物体認識に比べ，対象物の見た目が身体の姿勢に応じて変形し，色や模様なども同じカテゴリでも様々であることが挙げられる．例えば同じスカートという分類でも花柄のフレアスカートと黄色のマキシスカートでは見た目が大きく異なる．

衣服に関する研究で古くは上半身の衣服の形状の分類や [Borras 03]，変形する衣服の構造を And-Or グラフで記述するモデリング方法などが提案されていた [Chen 06].

画像認識に機械学習を用いた判別手法が一般化し，大規模な画像データセットを利用した研究が既に盛んになっていた 2012 年，著者らは変形する衣服形状をセグメンテーションによって認識する試みを発表した [Yamaguchi 12]. この研究では身体構造と衣服の着用部位に着目し，人物の姿勢認識を行なった結果を条件入力とする画像のセグ

∗1 Amazon Echo Look
∗2 fashionunited.com
∗3 https://deepanalytics.jp/compe/36
∗4 KDD 2017 workshop: Machine Learning meets Fashion
∗5 ICCV 2017 workshop: Computer Vision For Fashion

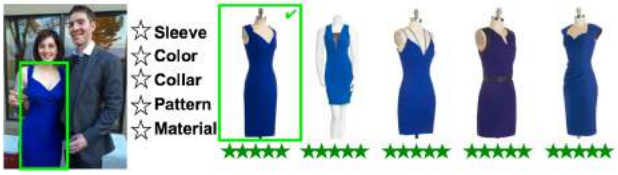


図 2: Street-to-shop 検索 [Kiapour 15]．左のスナップ写真に写る青いドレスと全く同一の商品を EC サイトから検索する．

メンテーション問題として定式化し，確率的推論によって解く手法を提案している．セグメンテーションはその後も様々な研究が提案されており [Dong 13, Liu 14a, Liu 14b, Simo-serra 14, Yang 14, Yamaguchi 15a]，特に近年では深層学習モデルを用いた高速で性能の高いセグメンテーションモデルが見られるようになっている [Liang 15b, Liang 15a, Tangseng 17].

2·2 Street-to-Shop とカタログ検索

画像を使ったファッション認識応用の中でも実用に結びつきが強いものが Street-to-Shop 検索だろう．これはユーザーがモバイル機器で撮影したストリートのスナップ写真から同じ EC サイトのショップ商品を検索するというもので，姿勢の推定を元に特徴量を抽出する手法 [Liu 12] や深層学習モデルを用いて類似度の計算を一貫して行う手法 [Kiapour 15] がこれまでにも提案されている（図 2）．また，EC サイトでのユーザー体験を向上させることを目的に，カタログの商品を形状などの細かい属性によってインタラクティブに並べ換える手法 [Kovashka 12] も試みられている．最近では属性の一部を書き換えた商品の検索を可能にするモデルも提案されている [Zhao 17].

2·3 スタイルの認識とリコメンデーション

個別の衣服の認識に対して全体の組み合わせ（Style, Outfit）の認識も試みられている．例えば [Simo-serra 15] では衣服の組み合わせやソーシャルメディアのプロファイル情報から Fashionability をスコアとして導き出し，ユーザーへの助言を行うパーソナルアシスタントを目指した手法を提案している．衣服の組み合わせ全体が醸し出すスタイルはそもそも明確な定義が難しく，教師データを用意することが難しい．そのため，著者らは [Kiapour 14] でスタイルの定義を言語によるタグ付けではなく，あるスタイルについて順位付けをしたデータを元にスタイルを構成する要素を分析するというアプローチを取っている．ファッションスタイルについて Yes/No で判別することが難しい場合でも，写真を見比べるとどちらがより Goth スタイルかといった判別を人は行いやすいためである [Parikh 11]．[Simo-Serra 16] では順位付けを持つ弱教師データから深層学習モデルによりファッションに適した特徴量表現を学習する手法を提案している．

2·4 ファッションの意味構造

衣服のカテゴリや色などの単純な視覚的要素とスタイルのような抽象化された要素の間にどのような関係があるのかといった問いに対し，近年ではデータ主導で分析する試みが見られるようになってきた．[Vaccaro 16] では画像を用いずに言語情報を用いて色などの低レベルな概念とスタイルのような高レベルな概念との間にどのような関係が成り立つのか，翻訳モデルを用いて相互の関係を分析している．

衣服の組み合わせに関しても，二つのアイテムがマッチするかという関係性をデータから学習をする試みが見られる．[Veit 15] では衣服のペアについて相性が良いか悪いかを，大規模な購買データを元に深層学習モデルで予測する手法を提案している．[Oramas 16] では衣服の組み合わせに関して衣服の部位の相互間の相性を分析し，組み合わせができるもの，できないものの関係をマイニングするアプローチを提案している．著者らは全身の衣服の組み合わせや属性全てを予測するモデルについて，CNN の予測結果から相互の関係性を同時確率により組み込んだ予測手法を提案している [Yamaguchi 15b].

2·5 弱教師データからの属性認識

ファッション画像は一般画像認識と比べて EC サイトなどの実応用に直接的に結びついており，学術研究も実環境を想定したノイズのある Web データを用いた研究が行われてきた [Bossard 12]．学術用に利用できるデータセットも Web データをクロールした弱教師データ [Yamaguchi 15a, Simo-serra 15] が主体であるが，最近ではクラウドソーシングを活用して Web データに大規模なアトリビュートのアノテーションを施した DeepFashion データセットも公開され [Liu 16a]，深層学習モデルの学習に利用されている [Liu 16b].

明示的なアノテーションが行われていない弱教師データから，教師データを発見するというアプローチもこれまでに提案されている．[Berg 10] では商品データと一緒に表出するキーワードの可視性をスコアとして求めてキーワードが学習可能な概念であるか判別し，また画像中での位置も求めるという手法を提案している．著者らはこの考え方を発展させ，深層学習モデルの内部発火を用いて可視性を導く手法を提案した [Vittayakorn 16]（図 3）．

2·6 トレンド分析

アパレル業界ではファッションショーから始まるトレンドをいち早く取り入れつつ季節ごとの商品をデザインする．このトレンドの分析はファッションについて深い知識と経験を持つエキスパートが担ってきたが，これを機械学習によって再現する試みが始まっている．[Hidayati 14] では身体の各部位についてパッチ特徴量をクラスタリングすることでショー単位でのトレンドの分析を試みている．

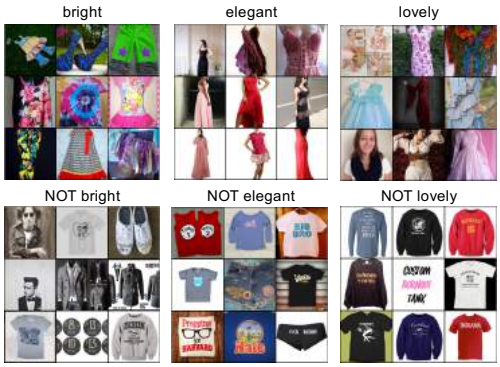


図 3: データから画像で判別可能なキーワードを検出 [Vittayakorn 16].

著者らは [Vittayakorn 15] で過去のファッションショーに現れたスタイルと，ソーシャルメディアに掲載されるストリートのスナップ写真との間の類似性を学習することで，特定のスタイルについてストリートでの人気の分析を行った．[Abe 17] ではワールドワイドのファッショントレンドの時系列分析を提案し，またそのための大規模なデータセットを収集している．最近ではファッショントレンドが将来どのように変遷するかを予測する手法も提案されている [Al-Halah 17].

2·7 ユーザー行動の理解

コンピュータビジョンを用いた視覚的な分析に加え，推薦などの実応用にはユーザー行動の理解も不可欠である．この方面から，ユーザーの行動データを用いたマッチするアイテムの予測と推薦手法がこれまでに提案されている [Veit 15, McAuley 15]．また，[He 16] ではユーザーの好みを反映したランキング関数を推定する手法を提案している．

人の行動データを取り入れた分析に関し，著者らはソーシャルメディア上でのファッション画像の人気度を予測するタスクに取り組んでいる [Yamaguchi 14]．具体的には，ファッションブログ Chictopia において画像やタグなどのコンテンツ要因や友人数などのソーシャル要因からユーザークリック数を回帰分析し，統計的にどのような要因がクリック数に影響しているのか分析した（図 4）．結果からは極めて強いソーシャル要因の影響が示唆されている．ファッション画像の品質については [Simo-serra 15] でも様々な要因を使った分析を行なっている．

2·8 スタイルの生成

明示的なアノテーションが存在しないデータから学習することができる深層学習手法として Generative Adversarial Networks (GAN) が注目を浴びているが，ファッション分野でも GAN をはじめとする生成モデルの研究が見られるようになっている．[Yoo 16] では衣服を着用したスナップ写真からカタログにあるような商品のみの画像

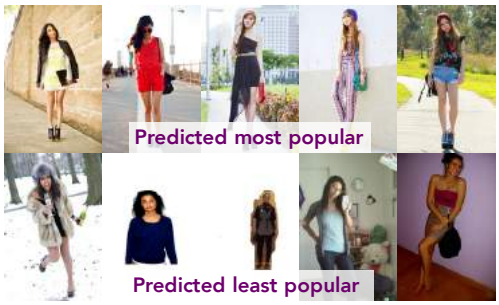


図 4: クリック数の予測 [Yamaguchi 14]．画像の品質やユーザープロファイルにより人気度は影響される．

を条件付き GAN により生成するという試みを報告している．[Lassner 17] ではセグメンテーション結果に条件付き Variational Auto Encoder でスタイル付けを行うという手法で衣服を着用した人物画像を生成している．また，[Ma 17] では姿勢を条件入力として異なる画像の人物を別の姿勢で再描画するモデルを提案している．スタイルの生成はヴァーチャルフィッティングに直接応用できる技術であり，今後も高品質な画像の生成が提案されることが予想される．

3. 今後の展望

高性能な深層学習モデルの普及とそのための大規模学習データの出現により，衣服の画像認識は既に商用を目指す段階に来ている．ファッション応用においても，どのように画像認識を解決するかという問題から，高度な画像認識を用いてどのような新たな問題を解き価値を生み出すかという点に研究の主軸が移ると予想される．

ファッション応用について，コンピュータビジョンや機械学習は言葉に明示できない概念を計量するツールとして機能する．例えば「フォーマルさ」のような概念を言語によって定義し，手動で画像特徴量を設計して画一的な基準で認識することはほとんど不可能である．データ主導のファッション分析により，このような定義の難しい概念をデータから統計的に学習することで定量化することが可能となった．また，定量化された概念を他の分野の問題に適用することが可能になり，例えばファッションに関するソーシャルメディアや EC サイトの分析，ファッショントレンドの定量分析などの応用が実現可能となった．見た目に関する，必ずしも言葉にできない概念さえも定量化，言語化こそが，ファッション分野にコンピュータビジョンがもたらす大きな利益である．

今後もコンピュータビジョン，機械学習を活用したファッションに関する新しい応用が研究が現れると想定されるが，視覚的な情報の定量化を含むような問題が成功するか否かを分ける鍵は学習データをどう設計するかという点が大きい．例えば，「原宿系」のような指標を学習する

には，Yes/No の絶対的な判断はデータの収集が困難であるが，二枚の画像の比較のような方法では判断がつく場合もある．また，あらゆるボキャブラリーを学習することは不可能であるが，どういった語彙が学習可能であるかデータから自動で判別し，語彙間の構造を理解するような手法がますます重要になってくる．これからも新しいファッション応用についてデータの収集から活用手法まで見据えた研究が進展することを期待する．

◇ 参 考 文 献 ◇

[Abe 17] Abe, K., Suzuki, T., Ueta, S., Nakamura, A., Satoh, Y., and Kataoka, H.: Changing Fashion Cultures, *arXiv* (2017)

[Al-Halah 17] Al-Halah, Z., Stiefelhagen, R., and Grauman, K.: Fashion Forward: Forecasting Visual Style in Fashion, *arXiv* (2017)

[Berg 10] Berg, T. L., Berg, A. C., and Shih, J.: Automatic attribute discovery and characterization from noisy web data, in *ECCV*, pp. 663–676, Springer-Verlag (2010)

[Borras 03] Borras, A., Tous, F., Lladós, J., and Vanrell, M.: High-level clothes description based on colour-texture and structural features, *Pattern Recognition and Image Analysis*, pp. 108–116 (2003)

[Bossard 12] Bossard, L., Dantone, M., Leistner, C., Wengert, C., Quack, T., and Van Gool, L.: Apparel classification with style, in *ACCV*, pp. 321–335Springer (2012)

[Chen 06] Chen, H., Xu, Z. J., Liu, Z. Q., and Zhu, S. C.: Composite templates for cloth modeling and sketching, *CVPR*, Vol. 1, pp. 943–950 (2006)

[Chen 12] Chen, H., Gallagher, A., and Girod, B.: Describing clothing by semantic attributes, *ECCV*, pp. 609–623 (2012)

[Chen 15] Chen, Q., Huang, J., Feris, R., Brown, L. M., Dong, J., and Yan, S.: Deep Domain Adaptation for Describing People Based on Fine-Grained Clothing Attributes, *CVPR*, pp. 5315–5324 (2015)

[Dong 13] Dong, J., Chen, Q., Xia, W., Huang, Z., and Yan, S.: A deformable mixture parsing model with parselets, *ICCV*, pp. 3408–3415 (2013)

[He 16] He, R. and McAuley, J.: Ups and Downs: Modeling the Visual Evolution of Fashion Trends with One-Class Collaborative Filtering, *WWW*, pp. 507–517 (2016)

[Hidayati 14] Hidayati, S. C., Hua, K.-L., Cheng, W.-H., and Sun, S.-W.: What are the fashion trends in new york?, in *ACM Multimedia*, pp. 197–200ACM (2014)

[Kiapour 14] Kiapour, M. H., Yamaguchi, K., Berg, A. C., and Berg, T. L.: Hipster wars: Discovering elements of fashion styles, *ECCV*, Vol. 8689 LNCS, No. PART 1, pp. 472–488 (2014)

[Kiapour 15] Kiapour, M. H., Han, X., Lazebnik, S., Berg, A. C., and Berg, T. L.: Where to Buy It: Matching Street Clothing Photos in Online Shops, *ICCV*, pp. 3343–3351 (2015)

[Kovashka 12] Kovashka, A., Parikh, D., and Grauman, K.: Whittlesearch: Image search with relative attribute feedback, in *CVPR*, pp. 2973–2980IEEE (2012)

[Lassner 17] Lassner, C., Pons-Moll, G., and Gehler, P. V.: A Generative Model of People in Clothing, *arXiv* (2017)

[Liang 15a] Liang, X., Liu, S., Shen, X., Yang, J., Liu, L., Dong, J., Lin, L., and Yan, S.: Deep Human Parsing with Active Template Regression, *TPAMI*, Vol. 37, No. 12, pp. 2402–2414 (2015)

[Liang 15b] Liang, X., Shen, X., Yang, J., Liu, S., Tang, J., Lin, L., and Yan, S.: Human Parsing with Contextualized Convolutional Neural Network, *ICCV* (2015)

[Liu 12] Liu, S., Song, Z., Liu, G., Xu, C., Lu, H., and Yan, S.: Street-to-shop: Cross-scenario clothing retrieval via parts alignment and auxiliary set, in *CVPR* (2012)

[Liu 14a] Liu, S., Feng, J., Domokos, C., Xu, H., Huang, J., Hu, Z., and Yan, S.: Fashion parsing with weak color-category labels, *IEEE Transactions on Multimedia*, Vol. 16, No. 1, pp. 253–265 (2014)

[Liu 14b] Liu, S., Liang, X., Liu, L., Lu, K., Lin, L., and Yan, S.: Fashion Parsing with Video Context, *ACM Multimedia*, Vol. 17, No. 8, pp. 467–476 (2014)

[Liu 16a] Liu, Z., Luo, P., Qiu, S., Wang, X., and Tang, X.: DeepFashion: Powering Robust Clothes Recognition and Retrieval With Rich Annotations, *CVPR*, pp. 1096–1104 (2016)

[Liu 16b] Liu, Z., Yan, S., Luo, P., Wang, X., and Tang, X.: Fashion Landmark Detection in the Wild, *ECCV* (2016)

[Ma 17] Ma, L., Jia, X., Sun, Q., Schiele, B., Tuytelaars, T., and Van Gool, L.: Pose Guided Person Image Generation, *arXiv* (2017)

[McAuley 15] McAuley, J., Targett, C., Shi, Q., and Hengel, A. v. d.: Image-based Recommendations on Styles and Substitutes, *SIGIR* (2015)

[Oramas 16] Oramas, J. and Tuytelaars, T.: Modeling Visual Compatibility through Hierarchical Mid-level Elements, *arXiv*, No. arXiv:1604.00036v1 (2016)

[Parikh 11] Parikh, D. and Grauman, K.: Relative attributes, in *ICCV*, pp. 503–510IEEE (2011)

[Simo-serra 14] Simo-serra, E., Fidler, S., Moreno-noguer, F., and Urtasun, R.: A High Performance CRF Model for Clothes Parsing, *ACCV*, pp. 2–11 (2014)

[Simo-serra 15] Simo-serra, E., Fidler, S., Moreno-noguer, F., Urtasun, R., and Rob, I. D.: Neuroaesthetics in Fashion : Modeling the Perception of Fashionability, *CVPR* (2015)

[Simo-Serra 16] Simo-Serra, E. and Ishikawa, H.: Fashion Style in 128 Floats: Joint Ranking and Classification Using Weak Data for Feature Extraction, *CVPR*, pp. 298–307 (2016)

[Tangseng 17] Tangseng, P., Wu, Z., and Yamaguchi, K.: Looking at Outfit to Parse Clothing, *arXiv* (2017)

[Vaccaro 16] Vaccaro, K., Shivakumar, S., Ding, Z., Karahalios, K., and Kumar, R.: The Elements of Fashion Style, *UIST* (2016)

[Veit 15] Veit, A., Kovacs, B., Bell, S., McAuley, J., Bala, K., and Belongie, S.: Learning Visual Clothing Style with Heterogeneous Dyadic Co-occurrences, *ICCV*, p. 27 (2015)

[Vittayakorn 15] Vittayakorn, S., Yamaguchi, K., Berg, A. C., and Berg, T. L.: Runway to realway: Visual analysis of fashion, *WACV*, pp. 951–958 (2015)

[Vittayakorn 16] Vittayakorn, S., Umeda, T., Murasaki, K., Sudo, K., Okatani, T., and Yamaguchi, K.: Automatic Attribute Discovery with Neural Activations, *ECCV* (2016)

[Yamaguchi 12] Yamaguchi, K., Kiapour, M. H., Ortiz, L. E., and Berg, T. L.: Parsing clothing in fashion photographs, in *CVPR*, pp. 3570–3577 (2012)

[Yamaguchi 14] Yamaguchi, K., Berg, T. L., and Ortiz, L. E.: Chic or Social: Visual Popularity Analysis in Online Fashion Networks, *ACM Multimedia*, pp. 773–776 (2014)

[Yamaguchi 15a] Yamaguchi, K., Kiapour, M. H., Ortiz, L. E., and Berg, T. L.: Retrieving similar styles to parse clothing, *TPAMI*, Vol. 37, No. 5, pp. 1028–1040 (2015)

[Yamaguchi 15b] Yamaguchi, K., Okatani, T., Sudo, K., Murasaki, K., and Taniguchi, Y.: Mix and Match: Joint Model for Clothing and Attribute Recognition., in *BMVC* (2015)

[Yang 14] Yang, W., Luo, P., and Lin, L.: Clothing co-parsing by joint image segmentation and labeling, *CVPR*, No. 2013, pp. 3182–3189 (2014)

[Yoo 16] Yoo, D., Kim, N., Park, S., Paek, A. S., and Kweon, I. S.: Pixel-Level Domain Transfer, *ECCV* (2016)

[Zhao 17] Zhao, B., Feng, J., Wu, X., and Yan, S.: Memory-Augmented Attribute Manipulation Networks for Interactive Fashion Search, *CVPR* (2017)

著 者 紹 介

山口 光太

2017 年 4 月入社．2017 年まで東北大学情報科学研究科助教．深層学習を用いた Web ビジュアルデータの分析研究に従事．2014 年 Stony Brook 大学コンピュータ科学の Ph.D. 取得．2008 年東京大学大学院情報理工学系研究科修士課程修了，2006 年東京大学工学部計数工学科卒業．2016 年，2015 年 MIRU 優秀賞受賞．IEEE，IPSJ，IEICE 会員．