

生成型深層学習による アパレルデザインのためのコーディネート画像生成

吉岡 朋友[†] 湯本 高行^{††} 山本 岳洋^{††} 大島 裕明^{†,††}

[†] 兵庫県立大学 応用情報科学研究科 〒650-0047 兵庫県神戸市中央区港島南町 7-1-28

^{††} 兵庫県立大学 社会情報科学部 〒651-2197 兵庫県神戸市西区学園西町 8-2-1

E-mail: [†]{aa20k512,ohshima}@ai.u-hyogo.ac.jp, ^{††}{yumoto,t.yamamoto}@sis.u-hyogo.ac.jp

あらまし 本研究では、アパレルデザインの制作を支援することを目的として、新しいコーディネート画像を生成する手法を提案する。パーソナライズされた服を購入しようとする、スーツなどのセミオーダーが主流で大抵は寸法を調整する程度のものである。ましてやデザインそのものを変更できるものは少ない。また、服のデザインは専門知識を持っていない一般人には難しい。そこで、本研究では機械学習を用いてコーディネート画像を生成する手法を提案する。この手法を用いて、一般人でも自分の理想とする新しいアパレルデザインを制作できるようにすることを目指す。ユーザは、「自分の理想とする服の形状のコーディネート画像」と「自分の理想とする服の柄のコーディネート画像」を入力することで、「理想の形状に対して、理想の柄を反映させたコーディネート画像」を作成する。具体的には、形状のコーディネート画像と柄のコーディネート画像を入力として受け取り、シルエット抽出を用いて画像を各パーツごとに分割し、個々のパーツごとに訓練した pix2pix モデルを用いることで入力された形状に対して柄を反映させた全身画像を出力する手法を提案する。

キーワード ファッション, 機械学習, デザイン, StyleTransfer

1 はじめに

現在、パーソナライズされた服を購入しようとする、スーツなどのセミオーダーが主流である。それは、各寸法を調整できるものがほとんどであり、デザインまで変更できるものは少ない。またデザインを変更できるものがあっても、せいぜい数パターンから選択するものであり、実際にパーソナライズされた服とは言えない。しかし、デザイナー以外の一般人が服の形状や柄を含めてデザインを行うことは難しい。現在の服の制作過程は、まずデザイナーが服のイメージであるデザイン画というものを作成する。次にその 2 次元のデザイン画をパタンナーと呼ばれる技術者が 3 次元のデザインに起こす。この一連の流れを商品ごとに行うために莫大な時間と労力を要する。

そこで、本研究では機械学習を用いてコーディネート画像を生成する手法を提案する。コーディネート画像とは、人の全身が写った画像のことである。図 1 は、コーディネート画像の例である。このように、実際の服を着た写真がコーディネート画像であるが、実際には存在しない服を着たコーディネート画像を生成することができれば、それをデザイン画の代替として十分に利用可能であると考えられる。

提案手法では、ユーザから 2 種類のコーディネート画像を入力として受け付ける。1 つは、自分の理想とする「形状」の服を着た人物が写ったコーディネート画像である。もう 1 つは、自分の理想とする「柄」の服を着た人物が写ったコーディネート画像である。提案手法が生成して出力するのは、理想の「形状」の服に対して、理想の「柄」を反映させたコーディネート



図 1 Fashion-Gen [14] データセット内のコーディネート画像の例

画像である。このような入出力を持つ手法を用いることにより、デザイン画を描けない人にも、既存には存在しない理想とするコーディネート画像を得ることができるようになると考えている。

服のデザインを構成する要素は「形状」、「柄」、「素材」の 3 要素であると言われている。この服のデザインを構成する 3 要素の中のうち、「素材」については、「形状」や「柄」に応じて服の作成時に考慮する必要がある要素である。そこで、本研究では「形状」と「柄」の 2 要素にターゲットを絞り、コーディネート画像生成を行うこととした。

図 2 は提案手法によってコーディネート画像を生成する流れ

の概略を表したものである。ユーザは、以下の2種類のコーディネート画像を入力する。

- 理想の「形状」のコーディネート画像
- 理想の「柄」のコーディネート画像

このとき、理想の出力は「形状」のコーディネート画像をベースとしており、その画像内に写った服に対して「柄」が反映されたようなコーディネート画像である。そのような理想の出力を得るために、提案手法では任意のコーディネート画像を入力として与えられた時に、その画像に対して理想の「柄」を反映させる機械学習モデルを構築する。そこでは、pix2pix [7] と呼ばれる技術を用いる。そのようなモデルが構築できれば、入力として与えられた理想の「形状」のコーディネート画像をそのモデルのその機械学習モデルに与えることによって、理想の出力が得られると考えられる。理想の「柄」を反映させる pix2pix のモデルを学習するためには、コーディネート画像と、その画像に理想の「柄」が反映されたコーディネート画像のペアが大量に必要となる。そのような画像ペアを生成するために、StyleTransfer [4] と呼ばれる技術を用いる。StyleTransfer とは、2枚の画像を入力とし片方の画像のスタイルをもう一方の画像に反映させる技術である。多くのコーディネート画像のそれぞれに対して、入力された理想の「柄」の画像のスタイルを反映させた新たなコーディネート画像を生成する。StyleTransfer によって作成したコーディネート画像とその画像に理想の「柄」を適用させた画像ペア集合を訓練データとして、pix2pix のモデル学習を行う。以上によって、ユーザの理想を反映させた新しいコーディネート画像の生成が行われる。

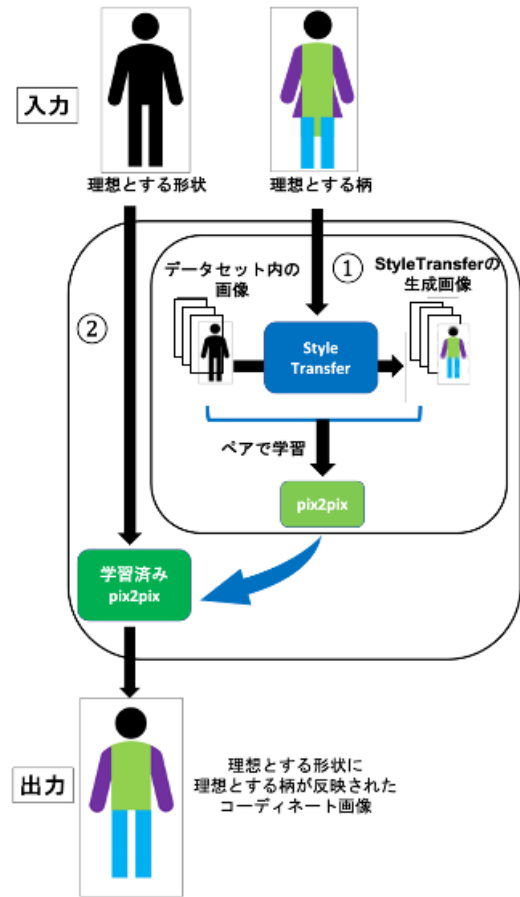


図2 コーディネート画像生成についての全体の流れ

2 関連研究

近年、機械学習によるデザイン生成は盛んであり、ファッション分野に関連した研究も同様である。

加藤ら [12] は敵対的生成ネットワークの1つである DCGAN [2] を用いて衣服をデザインした。この研究では 11,000 枚のファッションショーの画像を訓練データに用いて、画像を生成した。そして、その画像を元にしてパターンナーに依頼し実際に服を制作することまで行っている。

Zhu ら [16] は、人物画像と服のイメージを記述したテキストを入力とし、テキストに基づいた画像を生成する技術を提案した。この研究における訓練データはセンテンス付きの 79,000 枚の画像である。服の領域認識を行った後、形状生成の GAN と画像生成の GAN を組み合わせ、センテンスを反映した服のデザインを行った。

Yildirim ら [5] は、様々な服とポーズのモデルを高解像度で生成する技術を提案した。モデルとそのモデルが着用しているアイテムが含まれている 380,000 個の訓練データを用いて StyleGAN により画像を生成した。StyleGAN [17] とは、Progressive Growing GAN [10] で提案された手法を用いて GAN に比べて高解像度で精巧な画像を生成することが可能なモデルである。また、StyleTransfer [4] で提案された正規化手法である AdaIN [11] を用いて生成画像の制御を可能とした。モデル画像

には、そのモデルの姿勢を示すヒートマップが含まれており、そのマップを用いることで様々な姿勢のモデルを生成することができた。

Sbai ら [15] は、ファッション生成におけるデザインの新規性を高めるために異なる損失関数を持つデザイン生成のモデルを作成した。この研究の問題点として、まず始めにファッションデザインは、独自性がなければいけない。そして、生成されるデザインは必ず着ることのできるものでなければならないという点である。次に、GAN は訓練データの複製に特化しているが、その反面、独自性のあるデザインの生成は不得意という点である。4,157 枚の服の画像に対して、データオーギュメンテーションを行い、それを訓練データとして用いた。訓練データにはそれぞれ衣服とテクスチャのカテゴリがラベル付けされている。画像生成には、StyleGAN が用いられ、マスク化された衣服とテクスチャの情報を入力とし画像生成を行った。そして、生成画像に独自性を与えるために損失関数に Creative Adversarial Network(CAN) [1] で提案された損失関数を用いている。

Lee ら [8] は、アパレルデザイナーがどのように視覚的なインスピレーションを得て、デザインを作成するのか調査を行った。熟練のデザイナーに対して 6 種類の画像を提示した。次にその中から 1 枚の画像を選択してもらい、そこからインスピレーションを得てデザインを行った。デザイン制作過程の録画

とインタビューを通して、デザイナーのデザイン方法を明らかにした。選択された画像はノートルダム大聖堂であったが、デザイナーはそこに映る観光客や建物の窓枠などの詳細な部分を参考にし、デザインを行っていたことがわかった。そのために、画像生成技術を用いてデザイナーにインスピレーションを与えるためには、出力される画像は衣服のディテールなどを詳細に反映していなければならないと考える。

3 データと問題定義

本研究では、各ブランドのファッションショーの画像データ、ファッション SNS の一種である WEAR の画像データ、これらを用いてユーザが理想とするコーディネート画像を生成する問題に取り組む。本節では、本研究で用いる各種データとそのデータを用いたコーディネート画像生成についての問題定義を行う。

3.1 ブランドコレクションデータ

ブランドのファッションショーの画像を FASHION-SNAP.COM¹より収集した。

以下の 10 種類のブランドにおいて、それぞれ 1,000 枚を超える画像を取得した。

- COMME des GARÇONS
- Yohji Yamamoto
- Prada
- Balenciaga
- Sacai
- Toga
- Acne Studios
- VETEMENTS
- Alexander Wang
- Martine Margiela

3.2 ファッション SNS データ

WEAR²とは、ZOZO テクノロジーズが運営するファッションコーディネートに登録することができ、900 万件以上のコーディネートからファッションアイテムを探ることができるサービスである。登録ユーザが人気順に並べられたサイト内のページより、男女それぞれ上位 500 人の情報を取得した。さらに、そのユーザの中で登録コーディネート数が 100 を超えるユーザを対象に画像を取得した。理由としては、数枚しかコーディネート画像を投稿していないユーザの投稿画像は背景や被写体の写り方などの規則性がなく、訓練データとして用いるには不適であると判断したためである。対象ユーザは約 700 人で画像枚数は 100 万枚であった。

3.3 Fashion-GEN データセット

Rostamzadeh ら [14] により作成されたデータセットであり、293,008 枚のファッションアイテムの画像が含まれている。全

Long sleeve wool-blend coat in navy.
Peaked lapel collar. Double-breasted button closure at front.
Welt pocket at chest. Flap pockets and adjustable belt
with button fastening at waist. Central vent at back hem.
Full viscose satin lining. Black hardware.



図 3 Fashion-GEN データセット [14] 内のアイテムとそのアイテムについて記述されたテキストの例

てのアイテムにメインカテゴリとサブカテゴリが割り当てられている。ダウンジャケットを例にするとメインカテゴリは JACKET&COATS, サブカテゴリは DOWN である。またそれぞれのアイテムについて記述されたテキストを含んでいる。図 3 はアイテム画像とそのアイテムに対して記述されたテキストの例である。

3.4 問題定義

本研究の問題定義を行う。本研究で取り組む問題は「ユーザが理想とする服の形状の画像」と「ユーザが理想とする服の柄の画像」を入力し、「入力された服の形状と柄を反映したコーディネート画像」を生成することである。入力と出力の形は以下になる。

入力 服の形状の画像, 服の柄の画像

出力 入力された形状に柄を反映したコーディネート画像

3.5 評価手法

研究目的は「ユーザにより入力された服の形状の画像と柄の画像を反映した、ユーザが理想とするコーディネート画像を生成すること」であるために人手による評価を行う。

服の形状の画像と柄の画像を入力し同じ柄の画像だけで複数枚、出力する。その中に違う柄で学習された画像を 1 枚含めて被験者に提示し、その 1 枚を判別できるのか実験を行う。この実験において被験者が 1 枚だけ含まれる違う柄で学習されたコーディネート画像を判定することができれば、出力された画像はユーザが入力した柄をしっかりと反映されていると判断することができる。

次に既存のベースラインとの比較である。比較対象の画像生成のモデルは StyleTransfer である。1 つ目は、StyleTransfer に形状のコーディネート画像と柄のコーディネート画像を入力し、形状のコーディネート画像に対して柄のコーディネート画像の柄を反映して画像を出力する方法である。2 つ目は、StyleTransfer に前処理を行った画像を入力する方法である。ここでの前処理とは、入力された画像を身体のパーツに分けることである。身体の個々のパーツに対して StyleTransfer を用いて柄を反映させ、柄が反映された個々のパーツを最終的に組み合わせ、コーディネート画像を完成させる。上記 2 つの手法と

1 : FASHIONSAP.COM : <https://www.fashionsnap.com/>

2 : WEAR : <https://wear.jp/>

本研究で出力された画像を被験者に提示し、どちらの画像がより鮮明にかつ入力された柄を反映できているか判断してもらう。

4 提案手法

本節では、本研究で提案する手法について説明する。本研究の提案手法についての概要を述べた後に、個々に用いる生成技術についての詳細な説明を行う。

4.1 提案手法の流れ

ユーザの入力は「理想とする服の形状のコーディネート画像」と「理想とする服の柄のコーディネート画像」である。はじめにユーザの入力した柄のコーディネート画像に対して個々の身体のパーツに色分けされたセグメンテーション画像を取得する。次にそのセグメンテーション画像を用いて、個々の身体のパーツに分割された画像を取得する。また、訓練データに用いる WEAR の画像に対しても同様に個々の身体のパーツに分割した画像を取得する。

ユーザの入力した柄の画像の身体のパーツと訓練データの身体のパーツをそれぞれ対応する部位ごとで StyleTransfer を用いて訓練データに対して柄を反映させる。そして、柄が反映された個々の身体のパーツの画像を合成する。これにより、訓練データに対してユーザの入力した柄の画像を反映したコーディネート画像を生成する。

次に pix2pix の学習に移る。pix2pix には、背景が黒塗りにされた訓練データの画像と前段階で取得した柄の反映されたコーディネート画像を 2 枚 1 組で入力し、学習させる。

学習済みの pix2pix は、背景が黒塗りにされたコーディネート画像を入力するとはじめにユーザが入力した柄を反映したコーディネート画像を出力する。

4.2 WSHP を用いた入力コーディネート画像からのシルエット抽出とパーツごとのセグメンテーション

身体のシルエットを忠実に取得するために Fang ら [6] により提案された Weakly and Semi Supervised Human Body Part Parsing via Pose-Guided Knowledge Transfer (以下 WSHP) を用いてシルエット抽出を行う。WSHP は、姿勢がアノテーションされた画像を入力として受け取る。そして、入力された画像の姿勢と近い画像を人物の顔、胴体、腕、足のパーツがセグメンテーションされたデータセット内で複数枚探す。データセット内の複数のアノテーション画像との比較を行い、入力された画像の身体のセグメンテーションを行う。この処理を行うことで、入力された画像を人物の顔、胴体、腕、足の各パーツにセグメンテーションすることが可能となる。

4.3 StyleTransfer を用いた pix2pix の訓練データ作成

身体のパーツが個々に分割された画像を用いて StyleTransfer により、柄を反映させる。WSHP により身体のパーツを個々に認識し、個々のパーツのそれぞれに対して StyleTransfer を用いることで柄をより詳細に反映させることができると考える。そして、柄が反映された個々のパーツを組み合わせることでコーディネート画像とする。一連の流れを図 4 に示す。

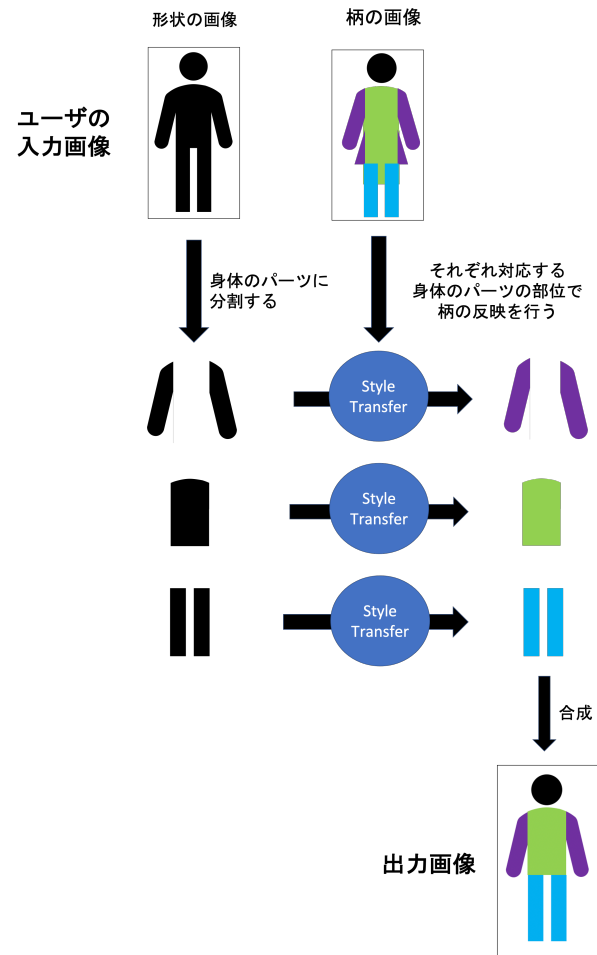


図 4 身体各パーツの組み合わせの模式図

ネット画像とする。一連の流れを図 4 に示す。

畳み込みニューラルネットワークの一種である StyleTransfer [4] は、構造を支配するコンテンツ画像と画風を支配するスタイル画像を入力とし、2つの画像の特徴を併せ持つ画像を生成するモデルである。具体的な生成方法については、コンテンツ画像とスタイル画像のそれぞれを訓練済みの VGG ネットワークに入力する。VGG ネットワーク [9] とは畳み込み層とプーリング層から構成される CNN であり、重みを含む層を 16 層または 19 層重ねたものである。ネットワークの特徴としては、小さいフィルターを持つ畳み込み層を 2 つから 4 つを連続して重ね、それをプーリング層を用いてサイズを半分にする。この流れを繰り返し行う。この VGG ネットワークを ImageNet [3] で学習されたものを StyleTransfer では用いている。ImageNet とは、1400 万枚以上の教師ラベル付き画像データベースである。その VGG ネットワークの途中の層で出力されたそれぞれの特徴ベクトルから損失関数を定義する。そして、その損失を最小化することを目的として学習を進めていく。これにより、2つの画像の特徴を併せ持つ画像を生成することができる。

身体の一部が分割された画像に対して、それぞれユーザが入力した柄を反映させる。この作業を全ての身体の一部に対して行う。

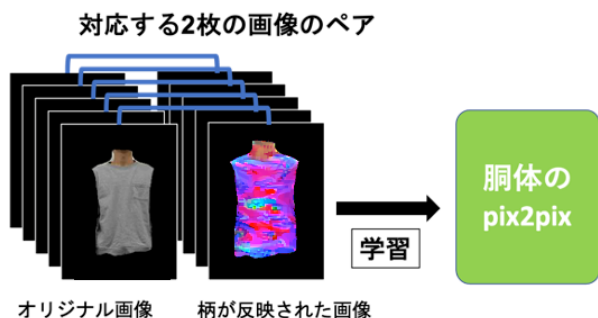


図 5 パーツ毎の pix2pix の学習に用いたペアの画像と学習の模式図

4.4 画像生成のための pix2pix を用いたシルエット画像からのデザイン生成

次に前段階で得られた StyleTransfer による生成結果の身体の部分画像と柄が反映される前の身体の部分画像をペアで pix2pix [7] を学習させる。図 5 にペアの画像で学習する際の模式図を示す。

pix2pix は GAN を用いた画像生成アルゴリズムの一種である。2つのペアの画像を用いてペア間の関係性を学習し、1枚の画像が入力されるとその1枚から関係性を考慮したもう一方のペア画像を生成する技術である。応用例としては、カラー画像とモノクロ画像をペアで学習し、モノクロ画像をカラー画像に変換するモデルがある。また、航空写真と地図をセットで学習し、航空写真から地図を生成するモデルがある。pix2pix に用いられる GAN は、cGAN (条件付き GAN) [13] である。通常の GAN は、Discriminator で入力されたデータが生成データなのかそれとも正解データなのか判定する。しかし、cGAN の Discriminator では正解ペアであるか不正解のペアであるかの判定を行う。また、PatchGAN [7] と呼ばれる手法を用いることで画像を小さなピクセル毎に真偽判定を行い、最終的にそれらを足し合わせることで全体の画像の真偽判定を行っている。

StyleTransfer によって生成された柄が反映されたパーツ画像と柄が反映される前の身体の部分画像のペアを訓練集合として pix2pix モデルを学習することで、図 6 に示すように、ユーザが理想とする形状のコーディネート画像を入力するとユーザが理想とした柄を反映した画像を出力するようになると考えられる。

5 まとめと今後の課題

本研究では、アパレルデザインの制作を支援することを目的として、新しいコーディネート画像を生成する手法を提案した。具体的には、ユーザが理想とする形状を表す画像とそれに適用したい柄を含むコーディネート画像を入力し、画像を各パーツごとに分割し、個々のパーツごとに訓練した pix2pix モデルを用いることで入力された形状に対して柄を反映させた全身画像を出力する手法を提案した。

今後は、実際にファッション SNS に投稿された画像および Fashion-GEN データセット中の画像に対して適用し、問題点

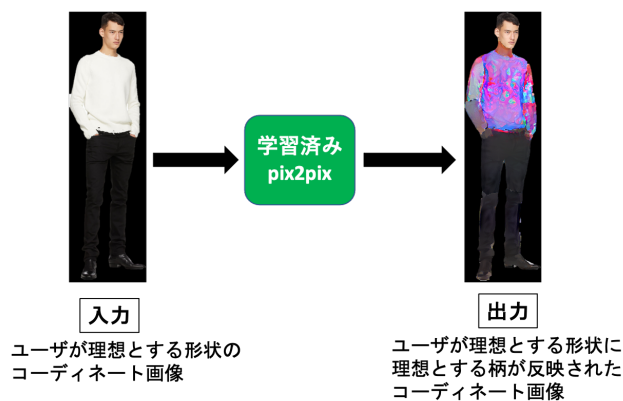


図 6 学習済み pix2pix の入力と出力の模式図

を発見し手法の改善を行っていく予定である。

謝 辞

本研究の一部は JSPS 科学研究費助成事業 JP18H03243, JP17H00762 による助成を受けたものです。ここに記して謝意を表します。

文 献

- [1] M. Elgammal Ahmed, Liu Bingchen, Elhoseiny Mohamed, and Mazzone Marian. CAN: creative adversarial networks, generating “art” by learning about styles and deviating from style norms. *arXiv:1706.07068*, 2017.
- [2] Radford Alec, Metz Luke, and Chintala Soumith. Unsupervised representation learning with deep convolutional generative adversarial networks. In *Proceedings of ICLR’16*, pp. 2414–2423, 2016.
- [3] Jia Deng, Wei Dong, Richard Socher, Li-Jia Li, Kai Li, and Li Fei-Fei. ImageNet: A large-scale hierarchical image database. In *Proceedings of CVPR’09*, pp. 248–255, 2009.
- [4] Leon A. Gatys, Alexander S. Ecker, and Matthias Bethge. Image style transfer using convolutional neural networks. In *Proceedings of CVPR’16*, pp. 2414–2423, 2016.
- [5] Yildirim Gokhan, Jetchev Nikolay, Vollgraf Roland, and Bergmann Urs. Generating high-resolution fashion model images wearing custom outfits. In *Proceedings of ICCV’19*, pp. 1–9, 2019.
- [6] Fang Hao-Shu, Lu Guansong, Fang Xiaolin, Xie Jianwen, Tai Yu-Wing, and Cewu. Lu. Weakly and semi supervised human body part parsing via pose-guided knowledge transfer. *arXiv:1805.04310*, 2018.
- [7] Phillip Isola, Jun-Yan Zhu, Tinghui Zhou, and Alexei A. Efros. Image-to-image translation with conditional adversarial networks. In *Proceedings of CVPR’17*, pp. 1125–1134, 2017.
- [8] Lee Jung, Soo and Jirousek Charlotte. The development of design ideas in the early apparel design process: a pilot study. In *Proceedings of CSCW’02*, pp. 126–135, 2015.
- [9] Simonyan Karen and Zisserman Andrew. Very deep convolutional networks for large-scale image recognition. *arXiv:1409.1556*, 2014.
- [10] Tero Karras, Timo Aila, Samuli Laine, and Jaakko Lehtinen. Progressive growing of GANs for improved quality, stability, and variation. In *Proceedings of ICLR’18*, 2018.
- [11] Tero Karras, Samuli Laine, and Timo Aila. A style-based generator architecture for generative adversarial networks. In *Proceedings of CVPR’19*, pp. 4401–4410, 2019.
- [12] Natumi Kato, Osone Osone, Hiroyuki, Daitetsu Sato, Naoya Muramatsu, and Yoichi Ochiai. Crowd sourcing clothes de-

sign directed by adversarial neural networks. In *Proceedings of NIPS'17*, 2017.

- [13] Mirza Mehdi and Osindero Simon. Conditional generative adversarial nets. *arXiv:1411.1784*, 2014.
- [14] Rostamzadeh Negar, Hosseini Seyedarian, Boquet Thomas, Stokowiec Wojciech, Zhang Ying, Jauvin Christian, and Pal Chris. Fashion-Gen: The generative fashion dataset and challenge. *arXiv:1806.08317*, 2018.
- [15] Sbai Othman, Elhoseiny Mohamed, Bordes Antoine, LeCuna Yann, and Couprie Camille. DeSIGN: Design inspiration from generative networks. *arXiv:1804.009021*, 2018.
- [16] Zhu Shizhan, Urtasun Raquel, Fidler Sanja, Lin Dahua, and Loy Chen Change. Be your own prada: Fashion synthesis with structural coherence. In *Proceedings of ICCV'17*, pp. 1680–1688, 2017.
- [17] Karrasand Tero, Laine Samuli, and Aila Timo. A style-based generator architecture for generative adversarial networks. In *Proceedings of CVPR'19*, pp. 4401–4410, 2019.