

LoRA, Stable Diffusion と EfficientNet を

用いたファッション・コーデイナー推勧モデルの開発

モウ スウシヨ

梶原 裕

有賀 爽玲

指導教員 劉 慶豊

要旨

人間が生活していく上で不可欠な基本的要素の一つが衣服である。正装があるように、衣服でその人物の印象、雰囲気大きく左右させる。このように服装は生活することにおいて大事な要素であるがゆえに、毎日のコーデに人は頭を悩ませる。そこで、一つのアイテムを決めるだけで、簡単にオシャレなコーデを提案してくれる機械を開発した。本研究では、2025年現在発信している、ランダムに選んだ各ブランドのコレクションのコーデ画像を学習させた LORA を用い、stable diffusion を用いて、画像生成を行った。さらに、stable diffusion からいくつか得られたコーデ画像の中からよりオシャレなコーデを選定するために、LORA の学習画像と同じ画像を学習した機械を用いて判定を行った。さらには、本研究で開発した機械を用いて、実際にアイテムの写真を撮り、そのアイテムを input としてオシャレ画像を生成してくれるようなアプリ開発も行った。この開発を行うことによって、専門的知識がない者でも簡単に機械を使えるように実用化を図った。

目次

第1章 序論

1.1 研究目的と背景

1.2 本論の構成

第2章 データの収集

2.1 各ブランドのコレクションからの収集

第3章 モデルについて

3.1 LORA の学習

3.2 Stable Diffusion を用いた着用画像の生成

3.3 EfficientNet を用いた画像のランキング

3.4 アプリケーション化

第4章 評価実験

4.1 実験タスク

4.2 実験環境

4.3 実験結果

第5章 考察

5.1 問題点

5.2 まとめ

参考文献

第1章 序論

1.1 研究目的と背景

衣服は人間が生活するうえで必ず関わるものである。卒業式や葬式などでは黒を基調にした衣服で悲しみを表し、結婚式や成人式などでは褐色の良い明るい衣服で喜びを表現する。就活など場面ではリクルートスーツというものがあるほどである。それほどどのような服装しているかで第一印象などが大きく変わる。この一般的に TPO(time, place, opportunity の略)と呼ばれるルールは、冠婚葬祭や会社での制服だけでなく普段の生活にも影響する。これからどこに行き何をするのかなどの状況によって服装を考えなければならない。また、自分をよく見せるためにもオシャレな服装を目指す必要がある。これは、服に興味のない者や、急いで家を出ていかなければならないときに頭を悩ませる問題の一つである。そこで、本研究ではこの問題を解決するために一つのアイテムを入力するだけでオシャレな服装を提案する AI を作成した。また、得られた画像がどのくらいオシャレであるかを評価するシステムを開発した。これらを組み合わせることで、複数枚出力された画像に対してどの程度オシャレであるかをそれぞれ点数化できる。その中で最も点数が高かったコーディネートをオシャレコーデとして提案できるのだ。また、本研究ではオシャレがしやすい冬のコーデ(冬のコーデは外気温が低いいため、重ね着(レイヤード)を必要とする。そのため、単純に使うアイテム数が多い、レイヤードの仕方によって印象が大きく異なるなどの要因がある。)の提案を目的としている。

1.2 本論の構成

第一章では、なぜこの研究をしようとしたかなどの研究目的や背景を記述した。第二章では機械学習をするうえで必要な学習用データの収集方法と出典について記述する。第三章では

実際に機会を構成している LORA や Stable Diffusion についての説明.また,生成した画像にランキングを付けるための EfficientNet を用いた画像分類モデルによるランキング手法を提案する.第四章では,この論文で提案している手法がどのくらいの精度を持っているかの検証をアンケートを用いて行い,結果を基に分析を行う.第五章では,ここまで得られた結果を基にした考察や,この研究で作成したプロジェクトの問題点や今後の展望について述べている.

第2章 データの収集

今回の研究では,オシャレなコーディネート画像を出力し,より良いコーデを提案することが目的である.それを実現するにはオシャレコーデをいくつも学習させなければならない.ここで問題となるのは,「オシャレ」とは何なのかである。「オシャレ」の感性は個人で大きく異なり,一般化するのは難しい.しかし,多くの人から「オシャレ」だと思われているコーデに対しては,ある程度の信憑性があると考えた.各ブランドのコレクションから持ってきたコーデを 413 枚取得し,それらを用いて学習を行った.また,本研究ではオシャレがしやすい季節である冬物のコーデの提案を目的としているため,データも冬のコーデのみを採用した.

2.1 各ブランドのコレクションからの収集

本研究では学習用画像を,FASHION PRESS という様々なブランドのコレクションをまとめているサイトを参照した.そのサイト内から現在立ち上げられているブランドのコレクションの画像をランダムに 413 枚収集した.これには,企業(各ブランド)が一般的に評価をされているコーディネーターに依頼しているため,一般的なオシャレに対して信憑性があると考えたためである.また,あるブランドから画像を多くとりすぎると,そのコーディネートをしたコーディネーターの趣向によって偏りによるバイアスが生じてしまうと考えた.そのため,1 ブランドに対して 20 枚まで収集すると制限を付け,収集者の恣意的な選択ではなく乱数を用いたブランドの選択を行うことによってバイアスを取り除いた.

第三章 モデルについて

3.1 LORA の学習

本研究では,画像生成の基盤モデルとして Stable Diffusion を採用し,ファッションコーディネートに関する知識を効率的に付加するために LoRA を用いた.

学習データには,2.1 節で作成したデータセットを用いた.

各画像に対して,BLIP 用いた自動タグ付けを行い,衣服の種類,色などに関するテキスト情報を付与した.(例:a man in a green coat and black pants) これにより,画像とテキストの対応関係を明らかにし,Stable Diffusion がコーディネートの学習に対し,可能な環境を構築した.

表 1

epoch	10
batch size	2
repeates	1
UNET lr	1.00E-04
text encoder lr	5.00E-05

LoRA の学習には,Kohya GUI を基にした Hollowstrawberry による LoRA Trainer を用いた.学習時の主な設定は,表 1 に示す.なお,画像解像度を 512×512 に固定して学習を行った.

3.2 Stable Diffusion を用いた着用画像の生成

本研究では,単一のファッションアイテム画像をクエリとして入力し,そのアイテムを用いた全身の着用画像を生成するために,Stable Diffusion に基づく画像生成モデルを用いる.具体的には,Stable Diffusion v1.5 を基盤とし,ControlNet を組み込んだ img2img 方式の生成パイプラインを採用することで,入力アイテムの外観的特徴を保持しつつ,多様なコー

ディネート候補を生成する.

生成の入力として, ユーザが与えたファッションアイテム画像を初期画像とし, CLIP により抽出されたアイテム記述および, 全身着用・写実性を強調するテキストプロンプトを同時に用いる. また, 入力ファッションアイテムの形を生成結果に反映させるため, LineartDetector により抽出した情報を入力として付与する. さらに, 不自然な人体表現や低品質な生成結果を抑制するため, ネガティブプロンプトを併用する.

本手法では, img2img 生成において比較的大きな strength 値を設定する. これにより, 入力アイテムの色や形状といった特徴を保持しつつ, 人物の姿勢, 背景, およびコーディネート全体の構成については一定の自由度を許容し, 多様な着用画像候補を生成することが可能となる.

さらに, ファッション領域に特化した生成を実現するため, 必要に応じて LoRA を導入する. LoRA を適用することで, 事前学習済みの Stable Diffusion モデルに対し, ファッション画像に特有のスタイルや分布を効率的に反映させることができる.

以上の手順により, 本研究では, 単一のファッションアイテム画像を入力として, 複数枚の全身着用画像候補を生成する. 生成段階では, 画像品質やファッション性にばらつきが含まれるため, 次節において識別モデルを用いたランキングを行う.

$$\text{pgen}(iq) = \{pk | pk = G(iq, c, l), k = 1, \dots, N\} \quad (1)$$

ここで, iq は入力されたファッションアイテム画像, c はテキストプロンプト等の条件, l は LineartDetector により得られた線画条件, $G(\cdot)$ は Stable Diffusion による生成モデルである.

3.3 EfficientNet を用いた画像のランキング

前節で生成された複数の着用画像候補の中から, よりファッション性の高いコーディネートを紹介するために, 本研究では EfficientNet を用いた画像分類モデルによるランキング手法を提案する.

本研究で用いる EfficientNet は, ファッションデータセット (class 1) とノーマルデータセット (class 0) の 2 クラス分類タスクとして学習されている. モデルは入力画像がどちらのクラスに属するかを予測し, softmax 関数により各クラスに属する確率を出力する. 本研究では, このうち class 1 に属する確率を「ファッションスコア」として定義する.

すなわち, 入力された着用画像がファッションデータセットに属する可能性が高いほど, その画像はよりファッション性の高いコーディネートであると解釈する. EfficientNet に入力する際には, 画像のリサイズおよび正規化を行い, 学習時と同一の前処理を適用する.

生成された各着用画像に対してファッションスコアを算出し, その値をランキングスコアとして用いることで, 生成画像群を降順に並べ替える. この手法は, 生成モデル自体の出力確率を用いてスコアリングを行う手法とは異なり, 外部の識別モデルによってファッション性を評価する点に特徴がある.

最終的に, 本研究では, ランキング結果の上位に位置する着用画像をユーザへの推薦結果と

して提示する。これにより、単に多様な着用画像を生成するだけでなく、ファッション性の観点から有用なコーディネートを選択的に提示することが可能となる。

$$p(x) = \text{softmax}(f_{\text{Eff}}(x)) = (p_0(x), p_1(x)) \quad (2)$$

ここで、 f_{Eff} は EfficientNet である、 $p_1(x)$ は生成着用画像 x がファッションデータセットに属する確率。

$$s(x) = p_1(x) \quad (3)$$

本研究では、 $p_1(x)$ をランキングスコア $s(x)$ として用いる。

$$x_{(1)}, x_{(2)}, \dots = \text{sort}_{x \in P_{\text{gen}}}(s(x)) \quad (4)$$

3.4 アプリケーション化

3.2 節および 3.3 節でそれぞれ構築した画像生成処理と画像ランキング処理の実行条件を統一し、専門知識がない者でも簡単に使えるように、Python と Streamlit を使用して WebApp 化を行なった。

本 WebApp は、ユーザーから入力されたファッションアイテム画像を起点として、画像生成処理および画像ランキング処理を 1 つのプロセス内で段階的に実行する構成とした。これにより、各処理間で中間結果をファイルとして保存・再読み込みする必要がなく、生成画像をそのまま次の処理へ受け渡すことが可能となっている。

またローカル環境で動作する構成にしつつ、外部からの動作確認およびデモ利用を目的として ngrok を用いた一時的な Web 公開を行った。ただし、本研究における ngrok の利用はあくまで検証およびデモを目的としたものであり、常時公開を前提とした運用は想定していない。

なお、本 WebApp は GPU 環境での段階的な処理を前提としており、複数の処理を同時に行う並列構成は採用していない。これは、画像生成処理が GPU 計算資源を集中的に消費するうえ、生成結果を入力として画像ランキングを行うため、処理を順次実行する必要があるためである。

さらに、計算資源の制約を考慮し、生成する候補画像の枚数や評価対象数には上限を設けている。これにより、処理時間の過度な増加を防ぎつつ、比較・評価が成立するだけの十分な数を確保している。

以上のように WebApp を使用することで、専門知識がない者でも簡単に使えるように実用化を図った。

第四章 評価実験

本手法の有用性を検証するために、評価実験を実施した。目的は以下の2点である。

1. 提案手法が、既存の生成手法および単純な生成結果の選択手法（ランダム選択、Stable Diffusion 単体）と比較して、入力されたファッションアイテムを適切に反映しつつ、おしやれさ、新規性、および着用意欲の観点において、より高い評価を得られるかを検証する。
2. 提案手法における生成画像の順位付け（EfficientNet による選択）が、100枚生成した画像からランダムに選択した場合と比較して、ユーザにとってより魅力的なコーディネート提案につながるかを明らかにする。

以上の目的のもと、ユーザの視点からコーディネート提案の有用性を評価するため、被験者として研究者の知人や親族、合計25名に協力してもらった。

4.1 実験タスク

本実験では、入力されたファッションアイテムに基づいてコーディネート画像を生成し、その推薦結果について評価を行った。被験者には、特定のファッションアイテム画像を入力とした場合に、各手法によって生成されたコーディネート画像を提示し、所定の評価項目に対して5段階リッカート尺度で評価値をつけてもらった。

4.2 実験環境

提案手法では、入力されたファッションアイテム画像を ControlNet および CLIP を用いて特徴抽出・条件付けし、Stable Diffusion と LoRA によりコーディネート画像を100枚生成。生成された画像に対して EfficientNet を用いて評価・順位付けを行い、上位1枚を最終的な推薦結果として被験者に提示した。

LoRA 学習用データセットの詳細は 3.1 節に示した通りである。また評価実験において入力として用いるファッションアイテム画像 ZOZOTOWN より収集した。なお収集した画像は10枚で tops, bottoms, shoes, hats, accessories の5カテゴリーにつき各2枚入っている。比較手法として、(1) 提案手法により生成した100枚の画像から無作為に選択した手法、(2) Stable Diffusion 単体による生成手法による手法を用いた。いずれの手法においても、被験者に提示する画像の形式および表示条件は統一した。

実装は Python を用いて行い、画像生成には Stable Diffusion v1.5 をベースとした派生モデルである Realistic Vision V5.1 を利用した。生成および評価処理は同一の実行環境上で行い、被験者への提示およびアンケート回答は Web ベースの評価インターフェースを通じて実施した。すべての被験者に対して同一の条件で実験を行うことで、手法間の公平な比較が可能となるよう配慮した。

4.3 実験結果

本実験では、図1のファッションアイテムを入力とし、4.2節に示す3種類の画像（提案手

法、ランダム抽出画像、Stable Diffusion 単体) について、被験者に比較評価してもらった。



図1 入力ファッションアイテム

以下に、被験者へ提示した質問内容を示す。

1. 画像内の人物のコーディネートはおしゃれと思うか？
2. ファッションアイテムは自然に映っているか？
3. 被写体は自然に映っているか？
4. 入力ファッションアイテム(入力画像)は生成画像内にどのくらい写っているか？

これらの質問に対して得られた評価結果を、図2に示す。

図2はリッカート尺度によって得られた評価の平均値を、図3は各評価の標本分散を示している。

ニット	1	2	3	4
Best	3.6	3.8	3.6	3.92
Random	3.6	3.96	4	3.88
Sdonly	2.48	2.88	2.8	1.68

革靴	1	2	3	4
Best	3.12	3.56	2.88	4.16
Random	3.96	3.84	3.16	1.68
Sdonly	3.44	3.72	3.76	1.48

ネックレス	1	2	3	4
Best	3.4	3.92	3.4	2.88
Random	2.96	3.68	3.32	2.64
Sdonly	2.24	3.04	3.08	1.48

ジップパーカー	1	2	3	4
Best	2.96	3.72	2.76	3.4
Random	2.92	4.08	2.92	3.84
Sdonly	3.4	3.68	4.16	1.96

スニーカー	1	2	3	4
Best	3.2	3.84	3.84	2.12
Random	3.04	3.84	3.96	2.12
Sdonly	1.88	2.68	3.56	1.52

サングラス	1	2	3	4
Best	2.8	3.44	3.48	3.8
Random	1.92	3	2.08	3.92
Sdonly	2.28	3.28	2.72	1.76

デニム	1	2	3	4
Best	3.28	3.56	3.08	3.52
Random	3.56	4	3.2	2.96
Sdonly	3	3.16	3.2	2.96

野球帽子	1	2	3	4
Best	3.6	3.56	3.12	2.64
Random	3.72	3.8	3.16	2.56
Sdonly	2.2	3.16	3.12	1.52

スラックス	1	2	3	4
Best	3.76	3.88	3.56	4.36
Random	3.8	3.96	3.6	4.08
Sdonly	3.48	3.48	3.92	2.12

青ニット	1	2	3	4
Best	2.48	3.76	3.24	3.84
Random	2.8	3.56	3.52	3.6
Sdonly	3.64	3.68	3.64	1.48

図2 リッカート尺度によって得られた評価の各平均値

ニット	1	2	3	4
Best	1.36	1.12	1.28	0.8736
Random	1.44	0.9984	0.72	0.9056
Sdonly	2.2496	1.9296	1.2	1.1776

靴	1	2	3	4
Best	1.1456	1.4464	1.5456	1.0944
Random	0.8384	1.2544	1.4144	1.5776
Sdonly	1.1264	1.4016	1.3024	0.7296

ネックレス	1	2	3	4
Best	1.44	0.7936	1.04	1.8656
Random	1.7984	1.2576	1.3376	1.5904
Sdonly	1.3824	1.9584	1.9536	0.7296

ジップパーカー	1	2	3	4
Best	0.7584	1.3216	1.5424	1.36
Random	1.1136	0.8336	1.6736	1.0144
Sdonly	1.92	1.0784	0.8544	1.3184

スニーカー	1	2	3	4
Best	1.68	1.0944	1.0144	1.7856
Random	1.9984	1.223333	1.0784	1.5456
Sdonly	1.3056	1.2256	1.5264	0.9696

サングラス	1	2	3	4
Best	1.44	1.5264	1.4496	1.28
Random	1.4336	2.24	2.3136	1.5936
Sdonly	1.6416	1.6416	2.1216	1.5424

デニム	1	2	3	4
Best	1.4016	1.3664	1.6736	1.5296
Random	1.2864	0.88	1.52	1.4304
Sdonly	1.6	1.8144	1.12	1.5104

野球帽子	1	2	3	4
Best	1.68	1.5264	1.6256	2.4704
Random	1.5616	1.36	1.3344	2.5664
Sdonly	1.84	1.6544	1.7056	1.0496

ストラックス	1	2	3	4
Best	1.3024	0.9856	1.6064	0.4704
Random	1.2	0.9184	1.52	0.8336
Sdonly	1.2096	1.2096	1.2736	1.8656

青ニット	1	2	3	4
Best	1.4496	0.8824	1.7024	1.4944
Random	1.84	1.0464	1.4496	1.52
Sdonly	1.1904	1.3376	1.3504	1.0496

図3 リッカート尺度によって得られた評価の各標本分散

図の見方について説明する。入力ファッションアイテムごとに表を作成しており、行は画像の種類を表す。Best は提案手法による画像、Random は出力画像からランダムに抽出した画像、Sdonly は Stable Diffusion 単体で生成した画像である。なお列は質問番号に対応している。

質問 1 (コーディネートの洗練度) : 本項目では、アイテム種別によって評価が分かれる傾向が見られた。ネックレス (3.40) やサングラス (2.80) , スニーカー (3.20) といった小物類を含む構成では提案手法 (Best) が他手法を上回った。一方で、ストラックスや野球帽子などの主衣服ではランダム抽出 (Random) が高い評価を得る傾向にあり、ジップパーカーや青ニットでは Stable Diffusion 単体 (Sdonly) が最高値を記録した。

質問 2 (ファッションアイテムの自然さ) : 全体として各手法間の差は僅差であり、いずれも高い自然性を維持していることが確認された。傾向として、小物類および青ニットを除く大半の項目で Random が最高評価を示したが、提案手法もネックレス (3.92) やサングラス (3.44) において同等以上の評価を得た。また、Sdonly は一部のアイテムで評価が低下したものの、ジップパーカー (3.96) など特定のアイテムでは他手法と同程度の評価を示した。

質問 3 (被写体の自然さ) : 人物および背景を含む被写体全体の自然性については、小物類において提案手法が最も高い評価を得た一方、衣類全般では Random または Sdonly が優位となる結果となった。特にジップパーカー (4.16) やストラックス (3.92) では Sdonly が最高値を得た。しかし、手法間のスコア差は概ね 0.5 ポイント以内にとどまっており、提案手法によって被写体の自然性が大きく低下する傾向は確認されなかった。

質問 4 (入力アイテムの再現性) : 本項目において、提案手法の有効性が最も顕著に現れた。Sdonly は全項目において著しく低い評価となり、入力アイテムの特徴保持が困難であることを示した。対して提案手法は、ジップパーカー、スニーカー、サングラスを除く全ての項目で最高評価を記録した。最高評価を逃した項目においても他手法との差は極めて僅少であり、多くのケースで入力画像の特徴を的確に生成画像へ反映できていることが実証された。特に革靴 (4.16) やストラックス (4.36) では 4 ポイントを超える高い再現性が確認された。質問 1.4~10.4 に対する評価について、生成手法 (Best, Random, Sdonly) を要因とした Friedman 検定を行った。その結果、すべての質問において手法間に有意な差が認められた ($\chi^2(2)=12.09-37.75, p<.01$)。事後検定として Wilcoxon の符号付順位検定を Bonferroni 補正付きで行った結果、多くの質問において Sdonly は Best および Random

より有意に低い評価を示した。一方、Best と Random の差は一部の質問を除き有意ではなかった。

第五章 考察

評価実験で得られた結果について考察する。研究目的にしていたオシャレなファッションコード生成において、1 の質問がその評価に値するが、提案手法の全てが一番高いポイントを得られるという結果ではなかった。しかし、4 の質問ではおおよそ全てのアイテムにおいて Best 画像の点数が最も高いため、入力アイテムを使いつつオシャレなコードを生成するという目的は達成しているのではないかと考えた。また、サングラスやネックレスなどの小物に関しての評価がどの質問に対して、Best 画像が他の画像よりも評価が高いことがわかる。これは、入力アイテムがどのコードに対してもマッチしやすいこと、画像を生成する際に被写体の体の一部を入力アイテムによって変更する箇所の面積が少ない、という理由が挙げられるのではないかと考えた。2,3 の質問に対してはどの画像も点数が高いことがわかる。これは、画像生成の観点において精度がよく、自然な映りになっていることが考えられる。実際、図 4 ようにどの画像も自然で映りがよいものとなっている。

ストラックスが入力アイテム
であった時の生成画像



図 4 生成画像の例

また図 3 に対して、質問 1 の列を見るとどのアイテムに対しての回答もある程度の散らばりがあることがわかる。これは、人によってオシャレに対する見方が違うことが大きな要因となっていると考えた。また、トップスやパンツに対しては Sdonly の分散が比較的小さく、ポイントも他の画像より小さいため、学習された LORA を用いたことが、オシャレな画像を作る糧になったとも考えられる。

5.1 問題点

前述の通り 2,3 の質問に対してはどの画像に対しても自然な映りであることが確認できた

が,中には全身の画像を出力するように命令しているものの,図5のように体の一部や入力アイテム自体を改変するだけの画像も多々存在した.



デニムのSdonly画像

革靴のRandom画像

サングラスのRandom画像

図5 全身を映し出せなかった画像

提案手法である Best 画像に対しては全身画像をすべてのアイテムで得られたが,全身のコーディネートを実際に得られるように改善が必要である.また,一部のアイテムにおいて入力アイテムの色や形を変えてコーディネートされている画像が何枚も確認できた.この入力アイテムの改変に対しても stablediffusion の改善が必要であると感じた.さらに,今回の画像を生成する際,時間がとてもかかってしまった.1.1でも述べたような実用化も視野に入れるのであるなら,生成時間については大幅に短縮する必要があると考えた.

5.2 まとめ

今回の研究では入力アイテムによって全身をオシャレにコーディネートする AI を開発した.研究途中では,参考資料の内容に加え,stablediffusion のバージョンを変えたり,使用する Lora の種類を変えたり,Clip などのよって入力画像を認識し,positive prompt にその情報を渡したりなど,様々な工夫を行った.しかし,5.1でも記述したように精度や生成時間などの観点においてまだまだチューニングを要するところが多々存在する.今後は本研究での問題点を払拭するとともに,より一般的なオシャレ画像を出力するように改善したいと思っている.そのためには,機械学習などの学習以外にもファッションなどの関連知識なども学習する必要があるのを感じた.また,院生とともに一つの研究に対して取り組むことによって,我々学部生はより高い知識や技術を学ばせてもらった.さらに,研究に対しての一連の流れを体験したことによってこれからの研究もやりやすくなるのではないかと感じた.この経験は院生,学部生問わず貴重なものであると感じた.

参考文献

- [1]大江 優真・荘司 慶行 (2025)「ファッションアイテム画像をクエリとする画像生成 AI を用いたコーディネート推薦」
- [2]株式会社 ZOZOTOWN(2021)
[画像ベースの仮想試着の実用化に向けた課題とアプローチ - ZOZO TECH BLOG](#)
- [3] Yiyang Xu, Wenjie Wang, Fuli Feng, Yunshan Ma, Jizhi Zhang, and Xiangnan He. Diffusion models for generative outfit recommendation. In *Proceedings of the 47th International ACM SIGIR Conference on Research and Development in Information Retrieval*, pp. 1350–1359, 2024.
- [4] Xintong Han, Zuxuan Wu, Yu-Gang Jiang, and Larry S. Davis. Learning fashion compatibility with bidirectional lstms. In *Proceedings of the 25th ACM International Conference on Multimedia*, MM '17, p. 1078–1086, New York, NY, USA, 2017. Association for Computing Machinery.
- [5] Yujuan Ding, P.Y. Mok, Yunshan Ma, and Yi Bin. Personalized fashion outfit generation with user coordination preference learning. *Inf. Process. Manage.*, Vol. 60, No. 5, sep 2023.
- [6] Wen Chen, Pipei Huang, Jiaming Xu, Xin Guo, Cheng Guo, Fei Sun, Chao Li, Andreas Pfadler, Huan Zhao, and Binqiang Zhao. Pog: personalized outfit generation for fashion recommendation at alibaba ifashion. In *Proceedings of the 25th ACM SIGKDD international conference on knowledge discovery & data mining*, pp. 2662–2670, 2019.
- [7] Xue Dong, Xueming Song, Fuli Feng, Peiguang Jing, Xin-Shun Xu, and Liqiang Nie. Personalized capsule wardrobe creation with garment and user modeling. In *Proceedings of the 27th ACM international conference on multimedia*, pp.302–310, 2019.
- [8] Rohan Sarkar, Navaneeth Bodla, Mariya I Vasileva, Yen-Liang Lin, Anurag Beniwal, Alan Lu, and Gerard Medioni. Outfittransformer: Learning outfit representations for fashion recommendation. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, pp. 3601–3609, 2023.

- [9] Si Liu, Jiashi Feng, Zheng Song, Tianzhu Zhang, Hanqing Lu, Changsheng Xu, and Shuicheng Yan. Hi, magic closet, tell me what to wear! In *Proceedings of the 20th ACM international conference on Multimedia*, pp. 619–628, 2012.
- [10] Tangwei Ye, Liang Hu, Qi Zhang, Zhong Yuan Lai, Usman Naseem, and Dora D Liu. Show me the best outfit for a certain scene: A scene-aware fashion recommender system. In *Proceedings of the ACM Web Conference 2023*, pp. 1172–1180, 2023.
- [11] Xiaoya Chong and Howard Leung. Body shape-aware object-level outfit completion for full-body portrait images. *Applied Sciences*, Vol. 13, No. 5, p. 3214, 2023.
- [12] Yashar Deldjoo, Zhankui He, Julian McAuley, Anton Korikov, Scott Sanner, Arnau Ramisa, Ren ´e Vidal, Maheswaran Sathiamoorthy, Atoosa Kasirzadeh, and Silvia Milano. A review of modern recommender systems using generative models (gen-recsys). In *Proceedings of the 30th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, pp. 6448–6458, 2024.
- [13] Maryam Moosaei, Yusan Lin, Ablaihan Akhazhanov, Huiyuan Chen, FeiWang, and Hao Yang. Outfitgan: Learning compatible items for generative fashion outfits. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 2273–2277, 2022.
- [14] Xintong Han, Zuxuan Wu, Zhe Wu, Ruichi Yu, and Larry S Davis. Viton: An image-based virtual try-on network. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 7543–7552, 2018.
- [15] Seunghwan Choi, Sunghyun Park, Minsoo Lee, and Jaegul Choo. Viton-hd: High-resolution virtual try-on via misalignment-aware normalization. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pp. 14131–14140, 2021.
- [16] Sen He, Yi-Zhe Song, and Tao Xiang. Style-based global appearance flow for virtual try-on. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 3470–3479, 2022.
- [17] Bochao Wang, Huabin Zheng, Xiaodan Liang, Yimin Chen, Liang Lin, and Meng Yang. Toward characteristicpreserving image-based virtual try-on network. In *Proceedings*

of the *European conference on computer vision (ECCV)*, pp. 589–604, 2018.

[18] Shizhan Zhu, Raquel Urtasun, Sanja Fidler, Dahua Lin, and Chen Change Loy. Be your own prada: Fashion synthesis with structural coherence. In *Proceedings of the IEEE international conference on computer vision*, pp. 1680–1688, 2017.

[19] Jonathan Ho, Ajay Jain, and Pieter Abbeel. Denoising diffusion probabilistic models. *Advances in neural information processing systems*, Vol. 33, pp. 6840–6851, 2020.

[20] Luyang Zhu, Dawei Yang, Tyler Zhu, Fitsum Reda, William Chan, Chitwan Saharia, Mohammad Norouzi, and Ira Kemelmacher-Shlizerman. Tryondiffusion: A tale of two unets. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 4606–4615, 2023.

[21] Yuhao Xu, Tao Gu, Weifeng Chen, and Chengcai Chen. Ootdiffusion: Outfitting fusion based latent diffusion for controllable virtual try-on. *arXiv preprint arXiv:2403.01779*, 2024. 1F-01 DEIM2025- 1F

[22] Jeongho Kim, Guojung Gu, Minho Park, Sunghyun Park, and Jaegul Choo. Stableviton: Learning semantic correspondence with latent diffusion model for virtual try-on. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 8176–8185, 2024.

[23] Aiyu Cui, Jay Mahajan, Viraj Shah, Preeti Gomathinayagam, Chang Liu, and Svetlana Lazebnik. Street tryon: Learning in-the-wild virtual try-on from unpaired person images. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 8235–8239, 2024.

[24] Robin Rombach, Andreas Blattmann, Dominik Lorenz, Patrick Esser, and Björn Ommer. High-resolution image synthesis with latent diffusion models. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pp. 10684–10695, 2022.

[25] J Redmon. You only look once: Unified, real-time object detection. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016.

[26] Alec Radford, Jong Wook Kim, Chris Hallacy, Aditya Ramesh, Gabriel Goh, Sandhini

Agarwal, Girish Sastry, Amanda Asbell, Pamela Mishkin, Jack Clark, et al. Learning transferable visual models from natural language supervision. In *International conference on machine learning*, pp. 8748–8763. PMLR, 2021.

[27] Pongsate Tangseng, Kota Yamaguchi, and Takayuki Okatani. Recommending outfits from personal closet. In *Proceedings of the IEEE International Conference on Computer Vision Workshops*, pp. 2275–2279, 2017.

[28] Edgar Simo-Serra, Sanja Fidler, Francesc Moreno-Noguer, and Raquel Urtasun. Neuroaesthetics in fashion: Modeling the perception of fashionability. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 869–877, 2015.

[29] Yuncheng Li, Liangliang Cao, Jiang Zhu, and Jiebo Luo. Mining fashion outfit composition using an end-to-end deep learning approach on set data. *IEEE Transactions on Multimedia*, Vol. 19, No. 8, pp. 1946–1955, 2017.

[30] Edward J Hu, yelong shen, Phillip Wallis, Zeyuan Allen-Zhu, Yuanzhi Li, Shean Wang, Lu Wang, and Weizhu Chen. LoRA: Low-rank adaptation of large language models. In *Proceedings of International Conference on Learning Representations*, 2022.

[31] Lvmin Zhang, Anyi Rao, and Maneesh Agrawala. Adding conditional control to text-to-image diffusion models. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pp. 3836–3847, 2023.¥

参考資料

株式会社カーリン (Carlin, Inc.) 「FASHION PRESS」

[2025-26 年秋冬コレクション一覧 - ファッションプレス](#)