

画像生成AIを用いた おしゃれなコーディネート 推薦モデル開発

データサイエンス&計量経済学研究室

23x4008 有賀 爽玲

23x4025 梶原 裕

25r6215 孟 巳初



Agenda

- 01 研究背景・目的
- 02 モデルについて
- 03 評価実験
- 04 結果
- 05 考察
- 06 結論
- 06 今後の展望



研究背景

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、一つのファッション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モデルを開発した。

従来手法

検索：持っているアイテムを使ったコーデが見つからない/見つかっても少ない

先行研究^{*1}：入力するファッションアイテムの特徴を十分に表現できない

先行研究^{*1}：不自然な背景や人物が生成されがち

先行研究^{*1}：ファッションアイテムに基づくランキングが有効であるとは言えない

提案手法

- 画像生成AIを用いることで持っているアイテムを使った沢山のコーデを提案できる
- ControlNet+CLIPを使う事でアイテムの特徴をキープしながらコーデを生成することが出来る
- LoRA+Realistic vision v5.1を使う事でよりリアルで自然な画像を生成できるように
- 数式ではなくおしゃれを学習したEfficientNetを用いることで安定してランキングを行えるように

^{*1} 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 (大江 優真 and 莊司 慶行,2025) 6 まとめと今後の課題 より抜粋

研究目的

本研究の目的は、単にStable Diffusionを用いて生成するだけでなく、おしゃれなコードを生成できることにある。そこで以下の3点を達成目標とする。

Target 01

モデルの構築

- ・画像生成AI(Stable Diffusion)を組み込み、検索よりも自由度の高いモデルを構築
- ・従来手法の問題点である入力アイテムを使用しない点、生成されたアイテム/被写体の不自然さ及びランキング手法の改善を行う

Target 02

生成画像の評価

- ・提案手法が、入力されたファッションアイテムを適切に反映しつつ、おしゃれさ・自然さ及び再現性の観点において、より高い評価を得られるかを検証する。
- ・提案手法における生成画像の順位付けが、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明らかにする。

Target 03

アプリ化

- ・専門知識がない者でも扱えるようにアプリケーション化を目指す
- ・段階としてはローカル環境で動くWebAPP, パブリックからアクセスできるWebAPP, スマホのアプリケーションの順に進めていく

モデル概要

01

入力

使いたいファッション
アイテムを入力します



02

提案手法

- ①入力アイテムを使った
全身コーデ100枚生成
- ②生成したコーデ画像
100枚にスコア付け、
最高評価のコーデを表示



03

出力

入力アイテムを使った
全身コーデ画像



①入力アイテムを使った全身コーデ100枚生成

提案手法について

- ①入力アイテムを使った全身コーデ100枚生成
- ②生成したコーデ画像100枚にスコア付け、最高評価のコーデを表示

ファッション
アイテムの入力



Phase 00

Phase 01

CLIPで入力アイテム
の概念を獲得。
画像生成AIに伝える

Stylish mustard yellow
knit sweater

ControlNet Lineartで
入力アイテムの形を保つよう
画像生成AIに制約をかける



Phase 02

Phase 03

画像生成AIで100枚生成
Stable Diffusion + LoRAで
生成する。



×100枚

②へ続く

①入力アイテムを使った全身コーデ100枚生成(詳細)

入力アイテムの概念を伝える CLIP

● アイテムの入力 + 辞書



Type&Color dict

Type={T-shirt, cardigan,
knit sweater, ...}
color={jet black, white,
wine red, olive green, ...}

● CLIP

“mustard yellow knit sweater”
と分類

● Stable Diffusionへ組み込む

SDの指示文(prompt)に組み込む.
...of a man wearing {**CLIP**}...の様に

入力アイテムの形を保つ ControlNet

● アイテムの入力



● ControlNet Lineart

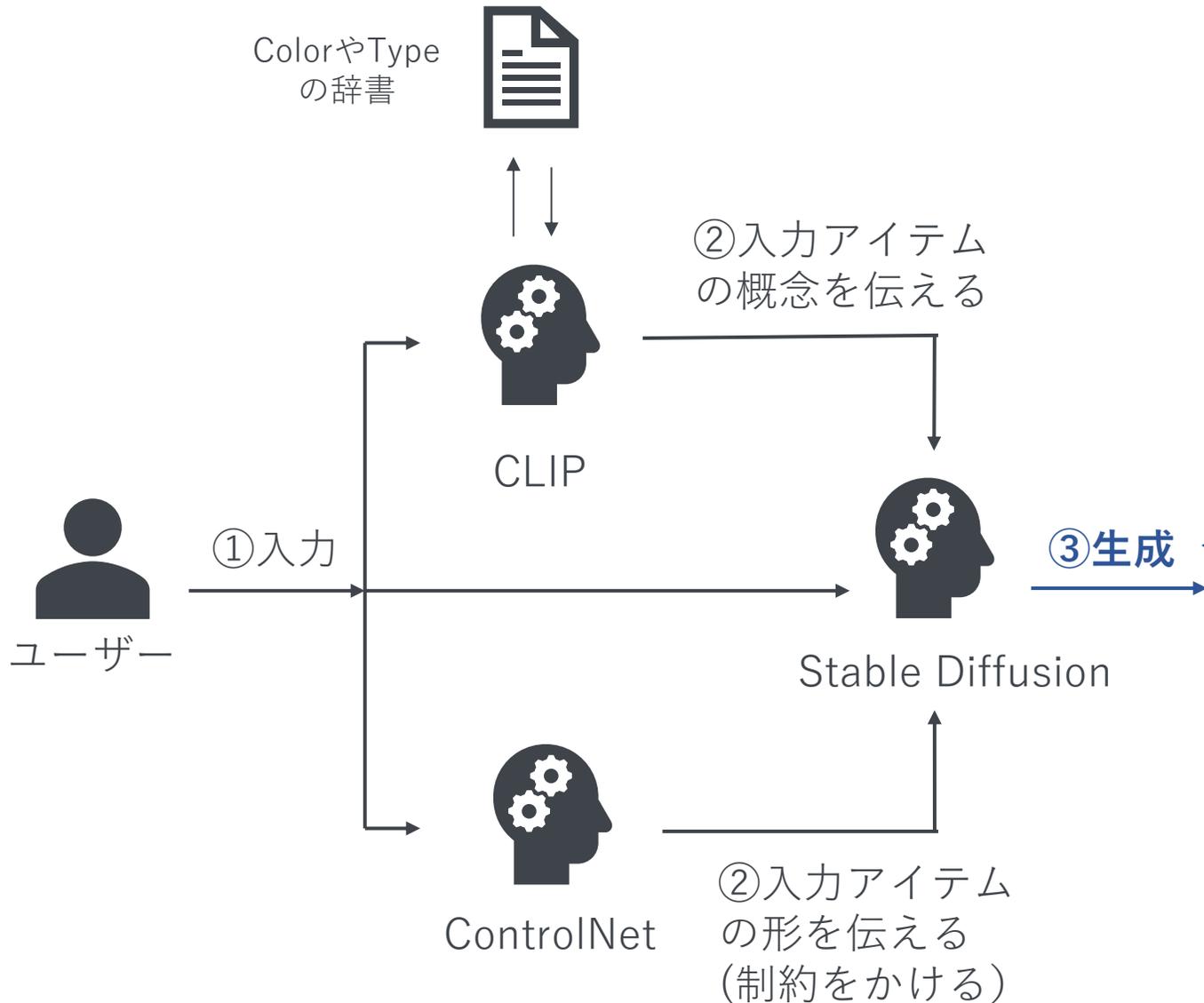
右図の線画が得られる.

図 ControlNet Lineart
で生成した線画

● Stable Diffusionへ制約をかける

SDにControlNetで得られた形をx%保持する
という制約をかける (今回は60%)

①入力アイテムを使った全身コーデ100枚生成(詳細)



生成について

・モデルはStable Diffusion SD1.5系の派生形**Realistic Vision V5.1**を採用。

・汎用的なモデルからファッションに特化させるため**LoRA***を採用。

*LoRAはファッションブランドからコーデ画像を400枚程度収集し、学習させた。



②生成したコードにスコア付け・最高評価のコードを表示

提案手法について

- ①入力アイテムを使った全身コーデ100枚生成
- ②生成したコーデ画像100枚にスコア付け。最高評価のコードを表示

①で生成したコーデ画像
100枚を入力



×100枚

Phase 04

Phase 05

rembgライブラリで
背景削除

背景によってスコアが
変わらないように
するため

EfficientNetでスコア付けを行う

1.000



0.998



...

Phase 06

Phase 07

最高評価のコードを表示



①から

②生成したコードにスコア付け・最高評価のコードを表示(詳細)

生成画像にスコア付けを行うEfficientNetについて

01 EfficientNetとは？

Mingxing Tan, Quoc V. Leは論文内*2で, called EfficientNets, which achieve much better accuracy and efficiency than previous ConvNets. と述べている. 次世代の画像認識モデル群.

*2:"EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks".arxiv.11 Sep 2020, <https://arxiv.org/abs/1905.11946>

02 データセット

WEAR*3及びFASHION PRESS*4で500枚ずつ収集. 写真の背景の削除・1000×1334へリサイズを行い, 写真の条件を揃えた.

*3:<https://wear.jp/>."その他のメンズのコーデ"

*4:<https://www.fashion-press.net/>."ファッションプレス: ファッションブランド・デザイナー情報"

03 訓練

WEARで収集した500枚を0(Normal)とし, FASHION PRESSで収集した500枚を1(Fashionable)として学習を行った. またデータ拡張を行った.

- ①Random Horizontal Flip : 50%の確率で左右反転
- ②Color Jitter : 明るさ・彩度・コントラストを±20%

04 モデルで使用



図 生成されたコーデ

生成されたコードに対して, 1(fashionable)~0(normal)の間でスコア付けを行う. 左図なら0.998

モデル概要

1

入力

使いたいファッション
アイテム

2

提案手法

画像生成AI+
ランキングモデル

3

出力

おしゃれなコーデ
の全身画像

2-1

CLIPで
アイテムの概念を伝える
+
ControlNet Lineart
でアイテムの形のキープ

2-2

Stable Diffusion
+
自作データセットで
学習済みのLoRAで
全身コーデ100枚生成

2-3

自作データセットで
学習済みの
EfficientNetで
スコア付けを行う

2-4

最高評価の
コーデを返す

評価実験

構築したモデル（提案手法）が良いものになっているか確かめるため、おしゃれさ/入力アイテム・被写体の自然さ/再現性の4観点から評価実験を行おうと考えた。

調査概要

調査名：

コーディネート生成画像に対する評価アンケート

調査方法：

Google formでの回答

調査期間：

2026/1/7~2026/1/8

回答者：

25名(min 20才, max48才)

属性：

男性19名(76%), 女性6名(24%)

評価尺度：

5段階のリッカート尺度

質問内容

質問1（おしゃれさ）：

画像内の人物のコーディネートはおしゃれと思うか？

質問2（自然性）：

ファッションアイテムは自然に映っているか？

質問3（自然性）：

被写体は自然に映っているか？

質問4（再現性）：

入力ファッションアイテム（入力画像）は生成画像内にとのくらい写っているか？

対象

種類：

Tops, Bottoms, Shoes, Hats, Accessories
の5種類. 1種類につき2枚で計10枚.これを以下の3手法で生成

比較手法：

比較用に,
提案手法/ランダム手法^{*5}/バニラ^{*6}
の3つで生成した画像を使用.

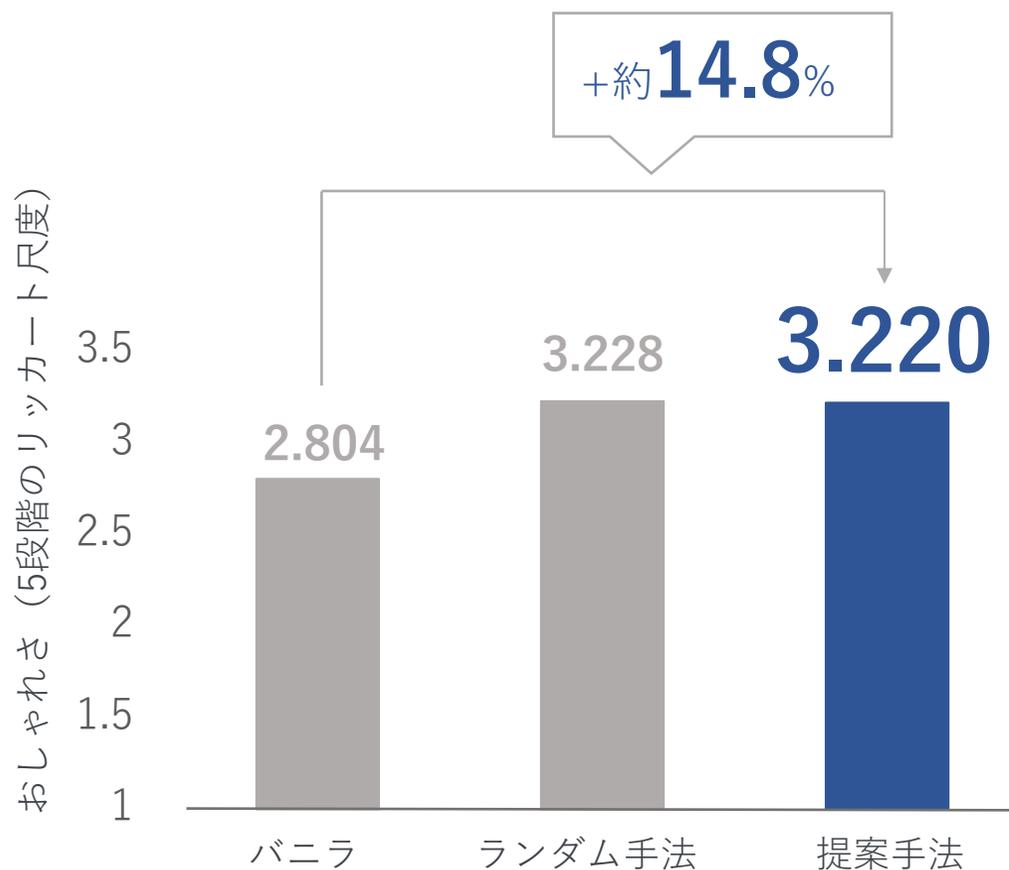
^{*5}ランダム手法→提案手法のEfficientNetを使用せずランダムに選択した画像

^{*6}バニラ→素のStable Diffusionのみで生成した画像（CLIPなし, ControlNetなし, EfficientNetなし）

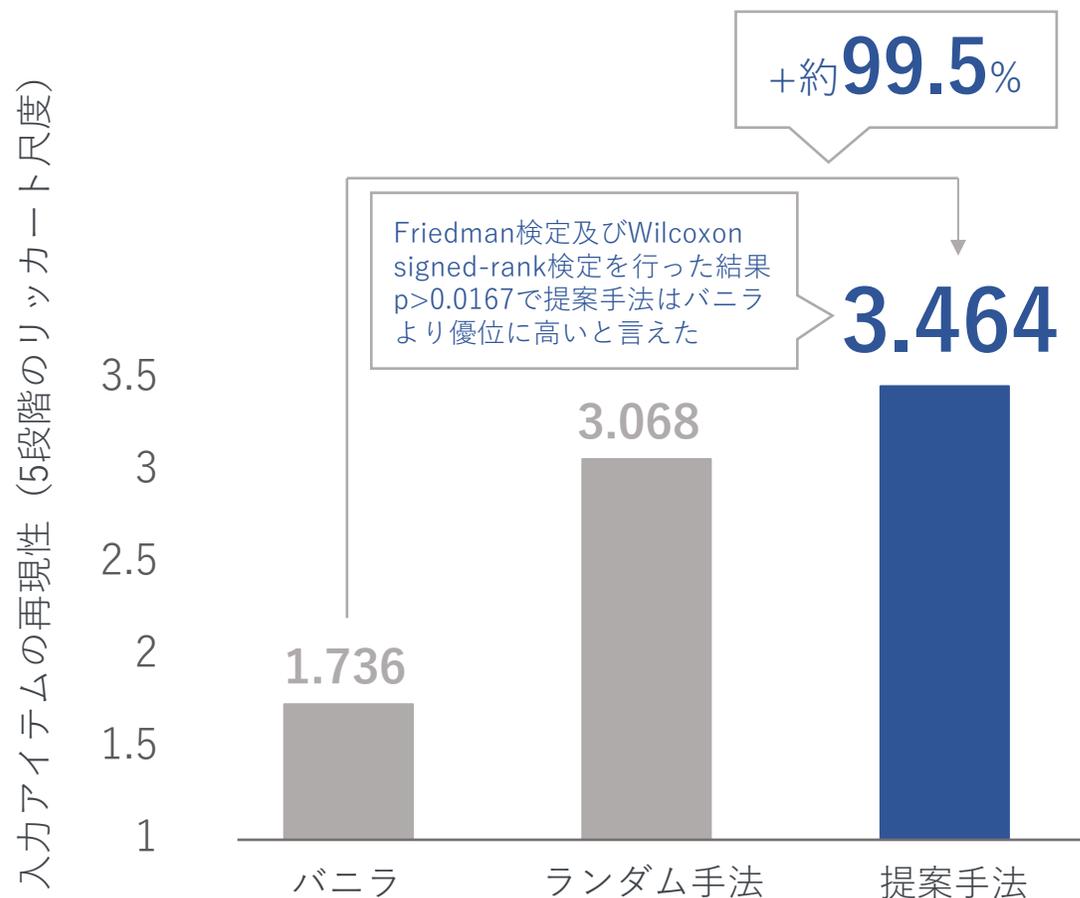
結果

おしゃれさでは、提案手法が何もしていないモデル（バニラ）に比べ、14.8%高い結果となった。入力アイテムの再現性は右図の通り。アイテム/被写体の自然性スコアは全手法同程度であった。

おしゃれさ



入力アイテムの再現性



考察①

考察① カテゴリ別にみていく。

Q1:

1. 提案手法がtops, accessoriesでは1位. 他3カテゴリではランダム手法について2位.

2. Accessoriesは提案手法と他手法で乖離あり

* 2.の深掘り

アイテムが小さいため上手くとらえきれず, ダサい/安っぽく見えるなどの画像になったのではないか

Q4:

1. 提案手法とランダム手法は、多くのケースで同程度の精度を維持

2. Shoesにおいてのみ、両手法の値に大きな差が発生

* 1.の深掘り

生成プロセスは同じであるためではないか

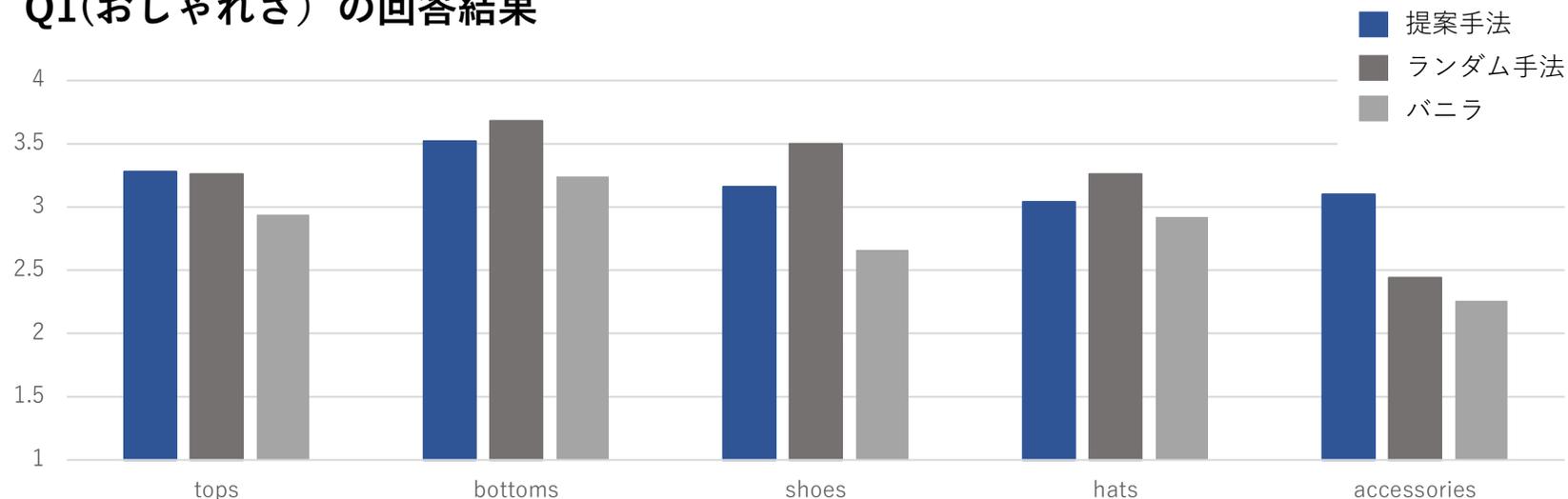
* 2.の深掘り

確認した所, ランダム手法の画像は見切れておりshoesが写っていなかったが提案手法は全身映っていた. EffNetが上手く作用したのではないか

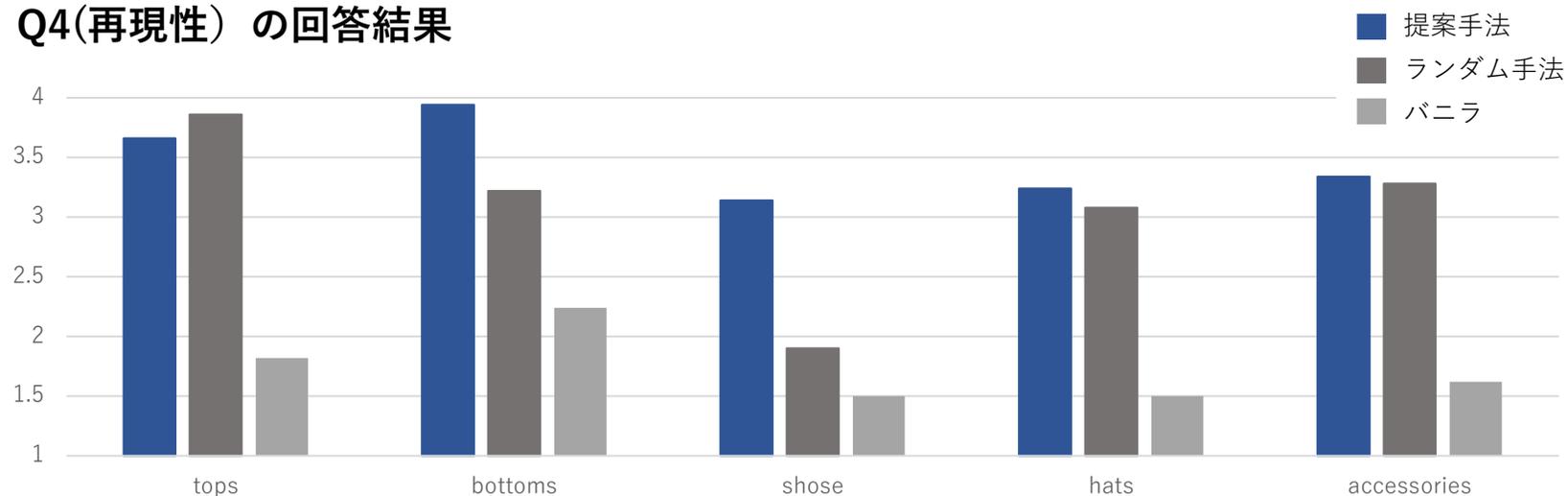
まとめ:

提案手法はおしゃれさで上位であり, 再現性でほぼ1位である事から, 入力アイテムを使いつつオシャレなコーデを生成するという目的は達成.

Q1(おしゃれさ) の回答結果



Q4(再現性) の回答結果



考察②

考察② なぜQ2,3はほぼ同じになったのか

Q2,3:

1. 提案手法と他手法との差はQ2で最大0.38, Q3では最大で0.1である.
2. 被写体の自然性はバニラが0.1提案手法よりも高い

* 1.の深掘り

結論, 元の生成モデルが全て同じもの(Realistic vision v5.1)を使用しているからではないだろうか.
右下の生成画像より全手法で安定して人の全身コーデが生成されていることが確認できる.

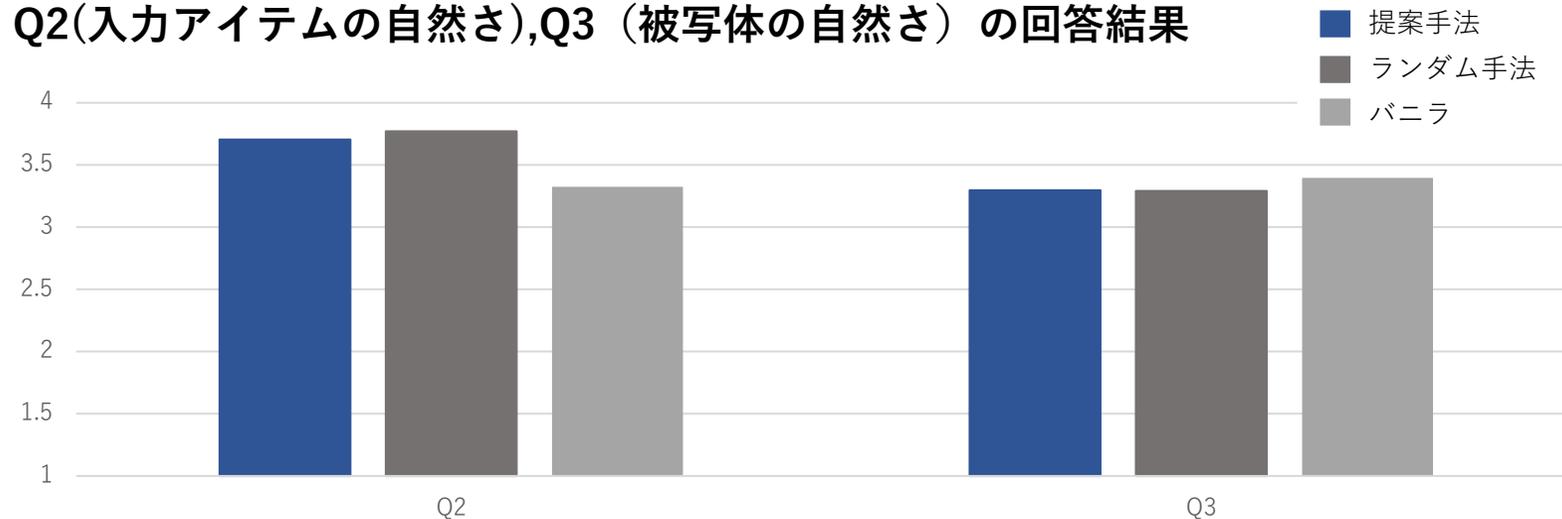
* 2.の深掘り

Friedman検定を行ったところ, p値が2.48e-01となり, 有意な差は認められなかった.

まとめ:

Stable DiffusionのRealistic vision v5.1を用いることで自然なコーデを安定して生成することが出来る.

Q2(入力アイテムの自然さ),Q3 (被写体の自然さ) の回答結果



生成画像

提案手法



ランダム手法



バニラ



考察③

考察③ おしゃれは人によって異なる

ヒートマップ：

1. どのアイテムに対しても回答のばらつきがあることが分かる
2. バニラの標準偏差が大きめ. topsでは1.5を超えている

* 1.の深掘り

おしゃれは人それぞれであることが分かる. そのため今後はユーザーにカスタマイズされたおしゃれを提案する方向性も考えられる.

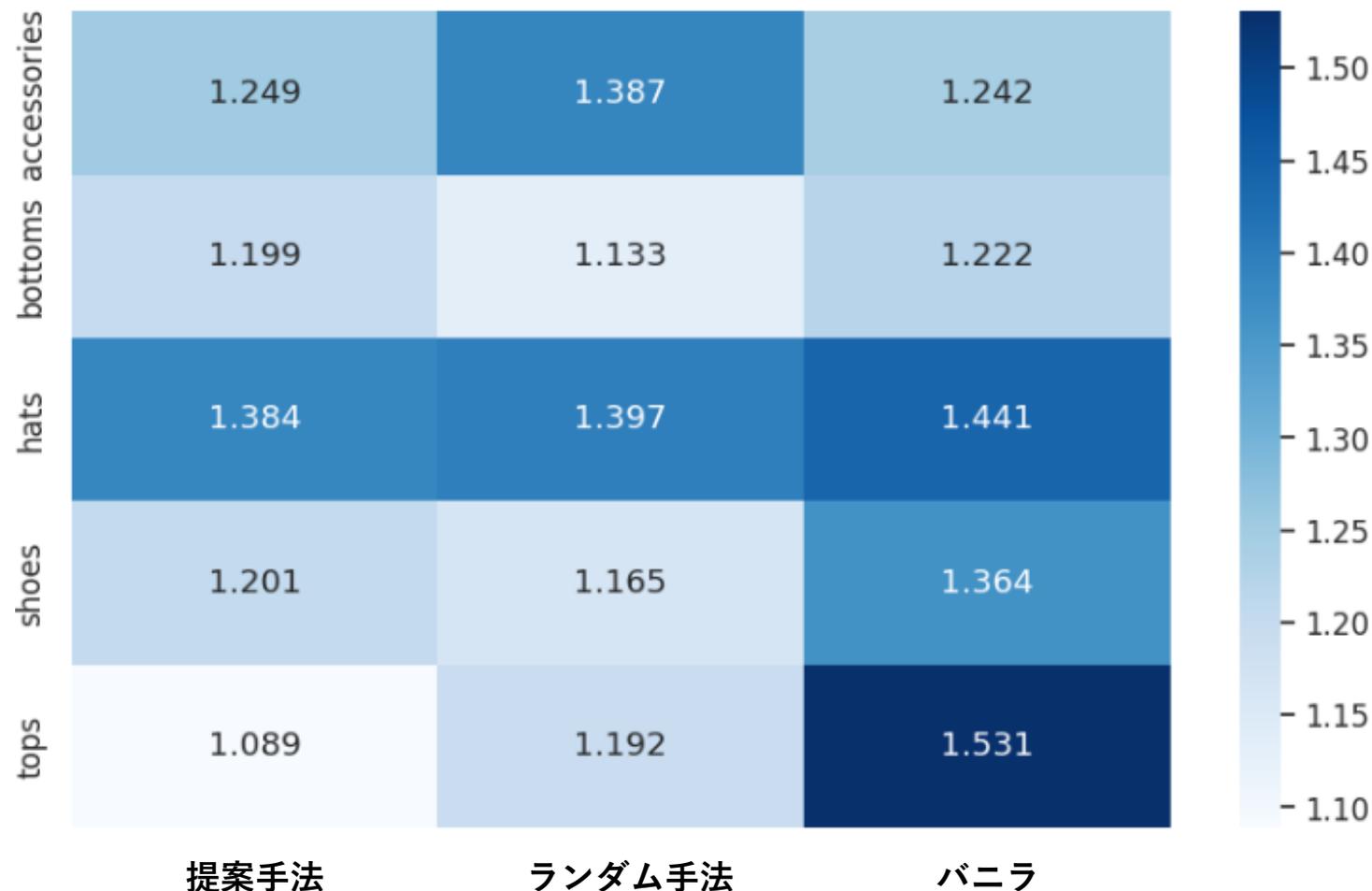
* 2.の深掘り

まずバニラの標準偏差が大きいことは好みも激しく分かれる. あるいは生成された画像の品質にムラがある可能性がある. topsではかなり奇抜なコーデとなっており意見が分かれたと考える. 反対に提案手法やランダム手法では1に近い. これはLoRAを用いることでおしゃれなものを生成する糧になったと考えられる.

まとめ：

おしゃれは人それぞれである.
提案手法はバニラに比べ標準偏差が小さい. これはLoRAの影響が考えられる.

Q1(おしゃれさ) の標準偏差のヒートマップ



WebAPP

Pythonで書き，StreamlitでWebAPP化，Ngrokで公開できるようにした。実際の画面は右図の通り。

CLIPによる
入力アイテムの
自動判定

EfficientNetによる
ランキングで
最高評価コードを
表示

GPU H100/80G
で提案まで

約4分

1. モデル初期化

LoRA Path

Classifier Path

モデル決定

AIファッション提案システム

2. アイテムをアップロード

アイテム画像をアップしてください

Drag and drop file here

Limit 200MB per file • JPG, PNG, JPEG

Browse files

test4.jpg
37.8KB



再解析

AI判定 大カテゴリ: clothing bottoms

アイテム詳細説明文 (変更可能, English only)

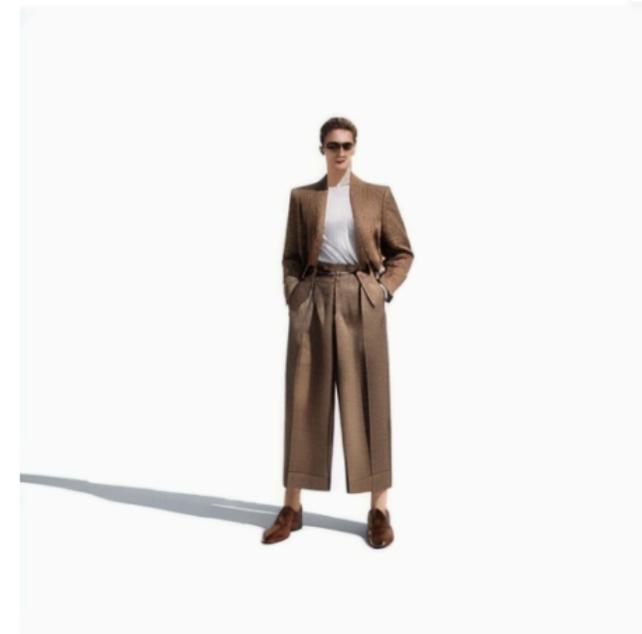
stylish brown slacks

コードを提案する

使い方ガイド

1. 左のサイドバーでモデルを初期化
2. 画像をアップロード (AIが自動解析)
3. 「コードを提案する」でAIが全身をスタイリング

おすすめのコーデ



Fashion Score: 1.000

本研究では、Stable Diffusionを中心に様々な手法を重ねておしゃれなコードを推薦するモデルを開発した。主な研究成果は以下の3点に集約される。

Result 01

モデルの開発に成功

- ・画像生成AI(Stable Diffusion)を組み込み、検索よりも自由度の高いモデルを開発
- ・評価の結果、入力アイテムを使いつつオシャレなコードを生成するモデルの開発に成功した

Result 02

モデルを評価し精度を確認

- ・モデルをおしゃれ/アイテム・被写体の自然さ/再現性の4項目で評価
- ・25名の評価結果より、今後への示唆を得ることが出来た。詳細は次スライドに記載する

Result 03

WebAPP開発に成功

- ・専門知識がない者でも扱えるWebAPP化に成功
- ・CLIPによって、アイテム名が分からなくても分類可能
解析ボタンを押して30秒で結果表示。間違えている場合、自分で修正することも可能*5

*5:評価実験では修正機能使用せず、CLIPの精度を見るため。

今後の展望

本研究で得られた知見と構築モデルに基づき、実務への応用と更なる発展のために以下の2つの展開を提案する。

機能の改善

全身のコーディネート
を確実に得られるよう改善

入力アイテム再現性
の安定化

生産時間の短縮



機能の追加

ユーザーの体型スタイルに
合わせたコーデ提案

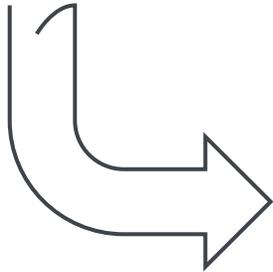
被写体をユーザーにして
コーデ画像を生成

ユーザー好みのファッション
スタイルに合わせたコーデ提案

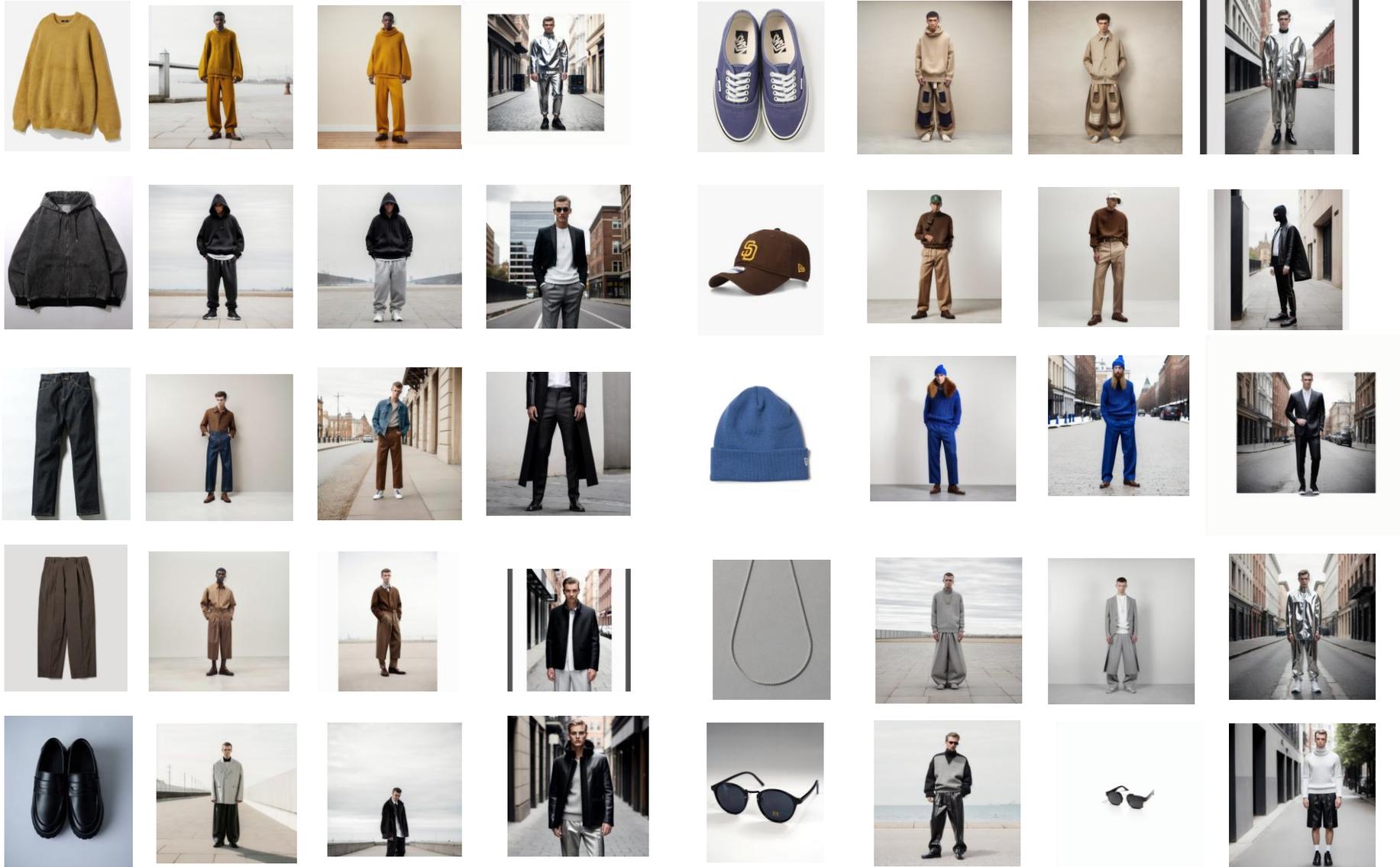
Appendix

おまけ：SD1.5系(提案手法)からFluxに変更し，生成してみた。

— Input画像



Appendix



(おまけ) : 生成画像

入力画像, 提案手法,
ランダム手法, バニラの
順番