# 服飾画像に対する印象推定のためのデータセットの構築

神戸 瑞樹[1,a)]　横山 想一郎[2,b)]　山下 倫央[2,c)]　川村 秀憲[2,d)]

**概要**：ファッション業界では感性が重要な地位を占めており，「かわいい」「ガーリー」といった感性を基に商品の開発や販売に関わる判断が下される．しかし，服飾に対する印象は受け手により変わり得るため，定量的な分析が出来ない．このため，人の印象を定量的に評価するシステムのニーズがある．本研究では，アノテーションのコストを一定としたときに画像枚数と一枚にタグ付けする人数を変化させ，深層学習を用いた推定結果を比較することで，印象を推定するためのデータセットの構築方法を検討した．

# Construct Dataset for Impression Estimation of Clothing Image

## 1. はじめに

ファッション業界では，「ガーリー」「上品」などの印象や各ブランドのイメージをもとに商品の開発・販売が行われている．こうした印象やイメージは明確な定義がなく，受け手により変わり得るものの，ファッション業界では一定の共通認識が存在する．しかし，定量的な分析が出来ず，デザイナーなどの一部の人によってどのような商品を作るかが決定されている．このため，人の印象を定量的に推定して，商品開発や売れ筋の分析などに役立てられるシステムが求められる．

著者らはこれまでの研究 [1] で，Fashion Impression Dataset (FID) を作成，学習し，1 枚の画像に 1 人分のタグが付けられた画像を大量に用意することで印象を学習できることを示した．しかしながら，FID でのデータの作り方が最適かは分からず，1 枚の画像に複数人でタグ付けしたデータを扱ったほうが効果的かもしれない．そこで，どのようにデータセットを作成すると効果的かを調査できるようなデータセットが必要となる．本研究では，1 枚の画像に複数人でタグ付けした画像を含む Fashion Multi-Impression Dataset (FMID) を作成し，画像枚数と 1 枚にタグ付けする人数を変化させて最適なバランスを調査する．

[1] 北海道大学大学院情報科学研究科
Graduate School of Information Science and Technology, Hokkaido University, Sapporo, Hokkaido, Japan
[2] 北海道大学大学院情報科学研究院
Faculty of Information Science and Technology, Hokkaido University, Sapporo, Hokkaido, Japan
[a)] kambe@complex.ist.hokudai.ac.jp
[b)] yokoyama@complex.ist.hokudai.ac.jp
[c)] tomohisa@complex.ist.hokudai.ac.jp
[d)] kawamura@complex.ist.hokudai.ac.jp

## 2. 関連研究

### 2.1 Fashion Dataset

AI の研究においてファッションはメジャーな分野の一つであり，ファッション用のデータセットが数多く作られている．特定のファッションカテゴリのみを含むもの [2] や多様なファッションアイテムを含むものが存在する [3]．より細かい分析を行うために属性やランドマークといった補足情報を含んでいるデータセットもある．この最大規模のデータセットとして DeepFashion があり，80 万枚の画像から構成されている [4]．これらのデータセットは，ほとんどが同一商品の画像検索を行うために設計されている．また，人の全身画像から個別のファッションアイテムを解析することを目的としたデータセットも存在する [5]．画像検索用のデータセットはイメージレベルでのアノテーションしか持っていないが，これらのデータセットはピクセルレベルでアノテーションを有している．一方で，現状のデータセットでは印象を取り扱っているものは少ない．Sirion Vittayakorn らが作成したデータセットには印象を表す属性も含んでいるが，視覚的なものが主となっている [6]．

### 2.2 Attribute Learning

属性を学習する研究はファッションに限らず行われている [7]．ファッションにおいては学習した属性を用いて，同一商品の画像検索 [4] やその属性を持っている服飾の検

© 1959 Information Processing Society of Japan

索 [8] などが行われている．また，WEB 上のノイズ混じりのデータから学習するものも存在する [6]．

これらの研究で扱う属性は基本的に視覚的なものであるが，印象的な属性を学習するものも存在する．Xiong らはサポートベクター回帰を用いて携帯電話の形状から印象を学習させた [9]．Vaccaro らは服飾画像に与えられたテキストから属性を抽出し，多言語トピックモデルを用いて視覚的な属性と印象的な属性の関係性を学習した [10]．Talebi らは CNN を用いて画像の「きれいさ」について学習させた [11]．

しかし，これらの印象的な属性を学習する研究では属性数・データ数といった面で規模が小さい．本研究では，属性数・画像枚数を大幅に増やしたデータセットを作成し，それを学習する．

## 3. Fashion Multi-Impression Dataset

FID から色違いを含まず 11000 枚を抽出した．簡単のため，抽出するものはトップスに絞った．付与するタグは FID から付与された回数が多いものの一部と 4 段階評価のものから選んだ．選んだタグは「おしゃれ」「きれい」「クール」「セクシー」「かわいい」「フェミニン」「モテ服」「デート」「カジュアル」「通勤・オフィス」の 10 個である．タグは 10 個しかないが，このデータセットでは，絶対的な正解がないデータを扱う際のデータの集め方を検証することを目的としているため，意見がばらばらになるタグを選べていれば，数は多くなくても十分であると考えたためである．タグは 4 段階評価 (1. 当てはまらない，2. どちらかというと当てはまらない，3. どちらかというと当てはまる，4. 当てはまる）で評価をしてもらった．4 段階評価にすることで，後でバイナリ評価に丸めることも出来るので，4 段階評価のまま学習した結果とバイナリ評価に丸めて学習した結果を比較できると考えたためである．ファッションの専門学校生 10 人にタグ付けを依頼し，10 人分のタグが付けられた画像を 2000 枚，4 人分のタグが付けられた画像を 1500 枚，2 人分のタグが付けられた画像を 2500 枚，1 人分のタグが付けられた画像を 5000 枚を作成した．10 人分のタグが付けられた画像には，それぞれの画像に 10 人全員がタグ付けを行った．4 人分のタグが付けられた画像には，10 人から 4 人を選ぶ組み合わせが 210 通りあるので，各組みに 7 枚ずつ割り振り，残りは 10 人がタグ付を行う回数だけを同じにしてランダムに割り振った．2 人分のタグが付けられた画像には，10 人から 2 人を選ぶ組み合わせが 45 通りあるので，各組みに 55 枚ずつ割り振り，残りは 10 人がタグ付を行う回数だけを同じにしてランダムに割り振った．1 人分のタグが付けられた画像には，10 人がタグ付を行う回数だけを同じにしてランダムに割り振った．最終的に 10 人がそれぞれ 3600 枚の画像に対してタグ付けを行った．**図 1** は画像と付けられたタグの分布の例である．

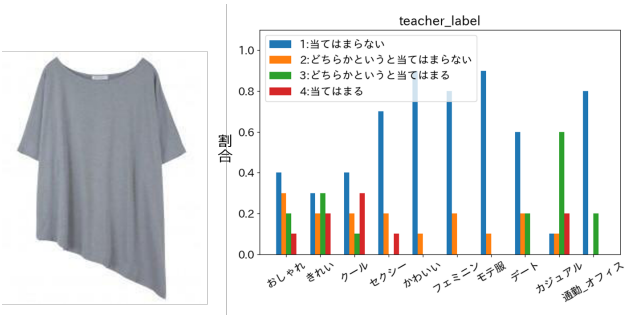



**図 1** FMID のサンプル
**Fig. 1** FMID sample

**表 1** 人ごとに 1,2,3,4 のそれぞれをつけた割合（1. 当てはまる 2. どちらかというと当てはまる 3. どちらかというと当てはまらない 4. 当てはまらない）
**Table 1** Percentage of 1, 2, 3, and 4 for each person (1.Strongly agree 2.Agree 3.Disagree 4.Strongly disagree).

| user_id | 1 の割合 | 2 の割合 | 3 の割合 | 4 の割合 |
|---|---|---|---|---|
| 1 | 0.753 | 0.017 | 0.109 | 0.121 |
| 2 | 0.654 | 0.079 | 0.113 | 0.153 |
| 3 | 0.617 | 0.123 | 0.165 | 0.095 |
| 4 | 0.349 | 0.113 | 0.266 | 0.272 |
| 5 | 0.510 | 0.258 | 0.192 | 0.040 |
| 6 | 0.546 | 0.113 | 0.164 | 0.178 |
| 7 | 0.427 | 0.165 | 0.209 | 0.198 |
| 8 | 0.735 | 0.072 | 0.160 | 0.033 |
| 9 | 0.392 | 0.156 | 0.224 | 0.228 |
| 10 | 0.717 | 0.116 | 0.093 | 0.074 |
| 平均 | 0.570 | 0.121 | 0.169 | 0.139 |

10 人分のタグが付けられた画像 2000 枚を対象として分析を行う．**表 1** は人ごとのタグ付けの傾向を示したものである．全体の傾向としては，当てはまらないを選択した割合が多く，人によってその度合が異なっている．**図 2**，**図 3** はそれぞれ「おしゃれ」「クール」に対する平均値のヒストグラム，分散のヒストグラムである．平均値のヒストグラムを見ると，全員が当てはまると言っているものはほとんどないが，全員が当てはまらないと言っているものは結構ある．ヒストグラムの形としては，山になっているものと，左上から下がっていく形になっているものがある．「おしゃれ」「きれい」「デート」「カジュアル」「通勤・オフィス」は山になっており，「クール」「セクシー」「かわいい」「フェミニン」「モテ服」は左上から下がっていく形になっている．平均値と分散のヒストグラムの形は似通っており，平均値のヒストグラムで山になっていれば，分散のヒストグラムでも山になっている．

### 3.1 Fashion Impression Dataset との比較

10 人全員がタグ付けをした画像を用いて比較を行う．**表 2** は 1 枚の画像に当てはまると答えたタグの平均個数を示している．4 段階評価を行ったものは 1,2 を当てはまら

© 1959 Information Processing Society of Japan

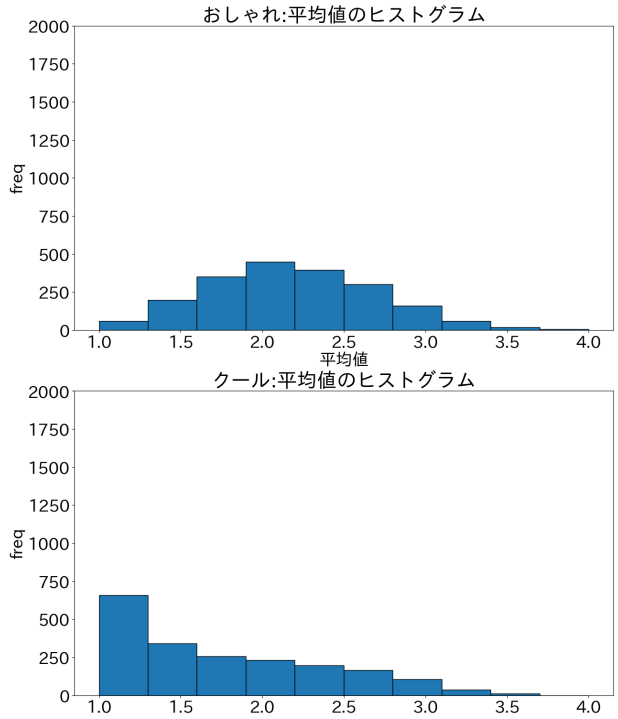



図 2 「おしゃれ」と「クール」に対する平均値のヒストグラム
**Fig. 2** Histogram of average value for ”fashionable” and ”cool.”

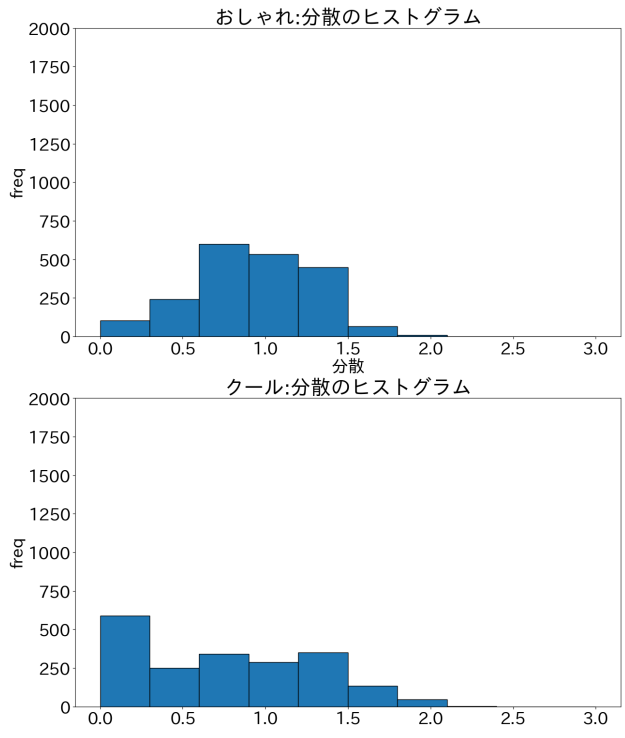



図 3 「おしゃれ」と「クール」に対する分散のヒストグラム
**Fig. 3** Histogram of variance for ”fashionable” and ”cool.”

表 2 一枚の画像に当てはまると答えたタグの平均個数．4 段階のタグは，当てはまる，どちらかというと当てはまるを当てはまるとして数えている．
**Table 2** Average number of tags that answered that they fit into one image. In four-stage tags, we treat strongly agree and agree as ”agree”.

| 対象 | 全てのタグ | 4 段階のタグ | バイナリのタグ |
|---|---|---|---|
| user1 | 2.288 | 1.454 | 0.716 |
| user2 | 2.681 | 1.385 | 1.079 |
| user3 | 2.608 | 1.178 | 1.367 |
| user4 | 5.426 | 3.612 | 1.429 |
| user5 | 2.371 | 1.467 | 0.748 |
| user6 | 3.44 | 1.683 | 1.537 |
| user7 | 4.126 | 2.397 | 1.309 |
| user8 | 1.943 | 1.138 | 0.774 |
| user9 | 4.566 | 2.371 | 1.813 |
| user10 | 1.692 | 0.91 | 0.666 |
| all user | 3.114 | 1.76 | 1.144 |
| FID | 4.253 | 3.043 | 1.209 |

表 3 FID で 4 段階評価だったタグにおける 1,2,3,4 を付けた割合
**Table 3** Percentage with 1,2,3,4 against four-stage tags in FID.

| | 1 の割合 | 2 の割合 | 3 の割合 | 4 の割合 |
|---|---|---|---|---|
| FMID | 0.589 | 0.128 | 0.17 | 0.113 |
| FID | 0.163 | 0.318 | 0.327 | 0.192 |

ない，3,4 を当てはまるとしている．これらを比較すると，FID で 4 段階評価であったタグの差が大きくなっている．これらの服に対して，4 段階評価であったタグの傾向を見ると，**表 3** のようになっている．FID では 2,3 を付けた数が多く，1,4 を付けた数が少なくなっており，FMID と傾向が大きく違うことが分かる．また，各服に対する FID と FMID の平均値の差を見ると，差が 1 付近になっている，即ち FMID で 10 人が一致した意見と真逆の意見を FID が持っている服も存在している．こうした差が生じた原因としては，コミュニティの違い，タグ付けを行った年の違いなどが考えられる．

## 4. Fashion Multi-Impression Dataset の学習方法

当てはまらない，どちらかというと当てはまらないを 0 に，どちらかというと当てはまる，当てはまるを 1 に変換したバイナリ評価での学習と，当てはまらないを 0 に，どちらかというと当てはまらないを 0.333 に，どちらかというと当てはまるを 0.667 に，当てはまるを 1 に変換した 4 段階評価での学習を行う．

学習を行うにあたっては，ResNet-50[12] を用いて，ImageNet で事前学習したものをファインチューニングしている．最適化手法としては AdamWR[13] を用い，バッチサイズは 128，weight decay は 0.0001，初期学習率は 0.0001

© 1959 Information Processing Society of Japan

表 4　あるタグに対して 1,2,3,4 のそれぞれをつけた人数
**Table 4**　Number of people who annotated 1, 2, 3, 4.

| | 1 | 2 | 3 | 4 |
|---|---|---|---|---|
| 服 A | 10 | 0 | 0 | 0 |
| 服 B | 0 | 0 | 0 | 10 |
| 服 C | 0 | 5 | 5 | 0 |
| 服 D | 10 | 0 | 0 | 0 |
| 服 E | 10 | 0 | 0 | 0 |
| 服 F | 0 | 5 | 5 | 0 |
| 服 G | 5 | 0 | 0 | 5 |

とした．

このデータセットでは，1 つの画像に対して複数の教師ラベルが与えられるが，学習時はこれらの平均を取らずにそれぞれ別のデータとして扱う．このため，1 エポックの学習内で同じ画像が複数回登場し，それぞれ違う教師ラベルを持つこととなる．

今回のデータセットは不均衡データとなっているため，損失関数に重み付けを行って対処する．ある服の可愛いに対して 10 人中 1 人が当てはまる，9 人が当てはまらないとタグ付けしていた時は，可愛いが 0.1 であると推測することが望ましい．しかしながら FID の時のようにラベルに対して重みを設定すると，データ全体の傾向としては可愛いと付けられた数の方が少ないため可愛いと付けられたデータに対する重みが大きくなる．このため，ロスが最小化する点では，0.1 よりも大きい数値を推測することになってしまう．これを避けるために，ラベルに対してではなく，服に対して重みを設定する．ここで，$dist_i$ は i 個目の分布の出現回数，$w_i$ を i 個目の分布に対する重みとし，$w_i dist_i = w_j dist_j$ となるように重みを設定する．よって，$w_i = \frac{\min_j(dist_j)}{dist_i}$ となる．

例えば，可愛いに対して**表 4** のようにタグ付けがなされていた場合を考える．出現回数が最も低い分布は服 B,G に対する分布で 1 回となり，服 A に対する分布は，A,D,E で 3 回出現しているので服 A,D,E に対する重みは $\frac{1}{3}$ となる．

2 クラス分類においては，服 F と G は分布は異なるものの，望ましい出力値はどちらも等しいため，これらは同じものとして扱って良い．即ち，2 クラス分類においては平均をとって，その値の出現回数に応じた重み付けを行うことになる．よって，服 F,G に対する重みは $\frac{1}{2}$ となる．

### 4.1　評価方法

10 人分のタグが付けられた画像 1000 枚を用いて評価を行う．正解ラベルを 0，0.333，0.667，1 に変換して平均をとり，その値と出力された値で MAE，相関係数を測る．また，その平均と出力値を比較し，これが± 0.1 以下なら正解として正解率を 0.25 区切りの区間で取り，それぞれの区間での正解率の平均 (平均区間正解率) を用いる．

## 5.　実験

### 5.1　印象推定のための効果的なデータ作成方法の調査

#### 5.1.1　目的

人の印象といった確固たる正解がないものを扱う際にどのようにデータを作るのが効果的かを調べる．タグ付けする人数と全部でタグ付け出来る回数が一定とした時，画像枚数と 1 枚にタグ付けする人数を変化させて，最適なバランスを調べる．

#### 5.1.2　手法

タグ付けを行う人数とタグ付けされる回数を一定とした時に，一枚にタグ付けする人数とタグ付けする画像の枚数を変化させてその結果を比較する．今回は 1000 枚の画像それぞれに 10 人がタグ付けしたものと，2500 枚の画像それぞれに 4 人がタグ付けしたもの，5000 枚の画像それぞれに 2 人がタグ付けしたもの，10000 枚の画像それぞれに 1 人がタグ付けしたものを比較する．10 人分のタグが付けられた画像の内 1000 枚はテストデータとすると，FMID は 10 人分のタグが付けられた画像 1000 枚，4 人分のタグが付けられた画像 1500 枚，2 人分のタグが付けられた画像 2500 枚，1 人分のタグが付けられた画像 5000 枚を訓練データとして持っている．10 人分のタグが付けられた画像から 4 人分のタグを抜き出すことで 4 人分のタグが付けられた画像 1000 枚を作成し，元からある 1500 枚と合わせることで，4 人分のタグが付けられた画像 2500 枚のデータを作成する．また，10 人分のタグが付けられた画像から 2 人分のタグを抜き出すことで 2 人分のタグが付けられた画像 1000 枚を作成し，4 人分のタグが付けられた画像から 2 人分のタグを抜き出すことで 2 人分のタグが付けられた画像 1500 枚を作成，元からある 2500 枚と合わせることで 2 人分のタグが付けられた画像 5000 枚のデータを作成する．同様にして 1 人分のタグが付けられた画像 10000 枚を作成する．抜き出す時は，元々付いていたタグからランダムに選択して抜き出しており，抜き出し方によって結果が変わってくるので，5 回平均で評価を行う．

学習のさせ方としては以下の 2 つを行う．当てはまらない，どちらかというと当てはまらないを 0 に，どちらかというと当てはまる，当てはまるを 1 に変換したバイナリ評価での学習．当てはまらないを 0 に，どちらかというと当てはまらないを 0.333 に，どちらかというと当てはまるを 0.667 に，当てはまるを 1 に変換した 4 段階評価での学習．これらそれぞれに対して一枚にタグ付けする人数とタグ付けする画像の枚数を変化させてその結果を比較する．

#### 5.1.3　結果・考察

**表 5** に学習結果を示す．相関係数，MAE について見ると，1 枚の画像に対してタグ付けしている人数が少ないデータに重み付けをするとかなり評価値が悪くなることがわか

© 1959 Information Processing Society of Japan

表 5　FMID の学習結果．訓練データの X*Y は Y 人分のタグが付けられた画像 X 枚を表す

**Table 5**　Learning result of FMID. X*Y in the training data represents X images tagged for Y people

| | | 訓練データ | 相関係数 | MAE | 平均区間正解率 |
|---|---|---|---|---|---|
| 重みなし | バイナリ | 1000*10 | 0.747 | 0.105 | 0.456 |
| | | 2500*4 | 0.749 | 0.106 | 0.481 |
| | | 5000*2 | 0.745 | 0.108 | 0.469 |
| | | 10000*1 | 0.736 | 0.11 | 0.456 |
| | 4 段階 | 1000*10 | 0.754 | 0.102 | 0.423 |
| | | 2500*4 | 0.754 | 0.102 | 0.438 |
| | | 5000*2 | 0.749 | 0.103 | 0.45 |
| | | 10000*1 | 0.735 | 0.106 | 0.429 |
| 重みあり | バイナリ | 1000*10 | 0.743 | 0.109 | 0.486 |
| | | 2500*4 | 0.726 | 0.137 | 0.467 |
| | | 5000*2 | 0.724 | 0.159 | 0.45 |
| | | 10000*1 | 0.722 | 0.183 | 0.404 |
| | 4 段階 | 1000*10 | 0.757 | 0.102 | 0.454 |
| | | 2500*4 | 0.723 | 0.115 | 0.471 |
| | | 5000*2 | 0.72 | 0.149 | 0.466 |
| | | 10000*1 | 0.715 | 0.177 | 0.426 |

る．これは，10 人中 1 人しか可愛いと言わない服があった時に，タグの抽出で可愛いと言っている方を抽出してしまうと，重みによって少数派の意見が多大な影響を持ってしまうためだと考えられる．**表 6** は正解ラベルの平均値によって区間ごとに区切ったものの正解率である．重み無しでは，1 枚に 10 人分のタグが付けられているものは，0.75˜1 における正解率が低い．これは，10 人分のタグが付けられていると，教師ラベルに大きい値が少なくなるが，1 枚に付けられるかタグの数が少ないとタグの抽出結果しだいで大きい値が教師ラベルに増えるためと考えられる．また，1 枚に 1 人分のタグが付けられているものは，中 2 つが低めになっている．これは，教師ラベルが 0,1 で表さられるためと考えられる．バイナリ評価と 4 段階評価を比べると，4 段階評価は 0.5 超えの正解率が低く，0.25˜0.5 が高くなっている．これは，ラベルを刻んだことで全体的に出力が真ん中によったためと考えられる．重みありでは，また，全体的に 0˜0.25 の精度が悪くなっており，重み付けによって出力が全体的に大きくなったためと考えられる．また，1 枚にタグ付けする人数が少なくなると精度が悪くなり，これは相関係数，MAE を見たときと同様の理由と考えられる．一方で 4 段階評価では，1 枚に 10 人がタグ付けしているデータの精度が低めになっている．これは，10 人の平均値をとった時にあり得る教師ラベルの値の数が多く意図したとおりに重み付けが効いていないためと考えられる．実際に，正解ラベルの値がほぼ同じでも，重みの値がかなり異なっている点があり，これを解消するためには，平均値が一定範囲のものの重みを同一にするといったことが考えられる．

**図 4** は 1000*10 の重み付きバイナリ評価における出力値例である．

表 6　FMID における区間ごとの正解率．A˜B は正解ラベルの平均値が A と B の区間内にあるものの正解率を表し，出力が正解ラベル± 0.1 以下なら正解としている．

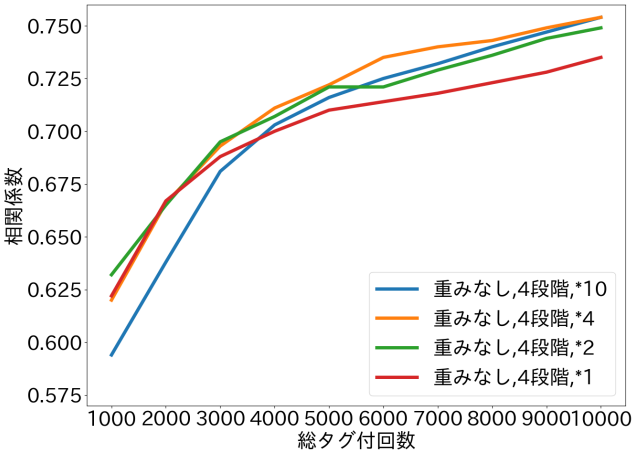
**Table 6**　Accuracy for each section in FMID. A textasciitilde B means the accuracy of the correct answer labels whose mean value is in the interval between A and B. If the output is within ± 0.1 of the correct answer label, the answer is correct.

| | | 訓練データ | 0˜0.25 | 0.25˜0.5 | 0.5˜0.75 | 0.75˜1 | 平均 |
|---|---|---|---|---|---|---|---|
| 重みなし | バイナリ | 1000*10 | 0.651 | 0.485 | 0.42 | 0.27 | 0.456 |
| | | 2500*4 | 0.657 | 0.472 | 0.429 | 0.366 | 0.481 |
| | | 5000*2 | 0.651 | 0.459 | 0.433 | 0.333 | 0.469 |
| | | 10000*1 | 0.667 | 0.434 | 0.386 | 0.338 | 0.456 |
| | 4 段階 | 1000*10 | 0.66 | 0.542 | 0.351 | 0.14 | 0.423 |
| | | 2500*4 | 0.678 | 0.512 | 0.371 | 0.19 | 0.438 |
| | | 5000*2 | 0.683 | 0.508 | 0.388 | 0.22 | 0.45 |
| | | 10000*1 | 0.691 | 0.473 | 0.338 | 0.212 | 0.429 |
| 重みあり | バイナリ | 1000*10 | 0.596 | 0.504 | 0.471 | 0.373 | 0.486 |
| | | 2500*4 | 0.516 | 0.407 | 0.398 | 0.548 | 0.467 |
| | | 5000*2 | 0.477 | 0.324 | 0.351 | 0.647 | 0.45 |
| | | 10000*1 | 0.394 | 0.273 | 0.336 | 0.612 | 0.404 |
| | 4 段階 | 1000*10 | 0.633 | 0.555 | 0.44 | 0.187 | 0.454 |
| | | 2500*4 | 0.593 | 0.496 | 0.437 | 0.357 | 0.471 |
| | | 5000*2 | 0.384 | 0.436 | 0.488 | 0.559 | 0.466 |
| | | 10000*1 | 0.23 | 0.361 | 0.564 | 0.55 | 0.426 |


図 4　1000*10 の重み付きバイナリ評価における出力値例

**Fig. 4**　Example of output value in 1000 * 10 weighted binary evaluation.

© 1959 Information Processing Society of Japan



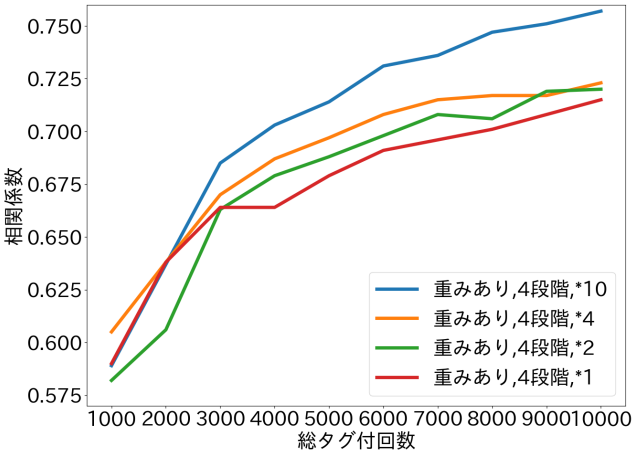
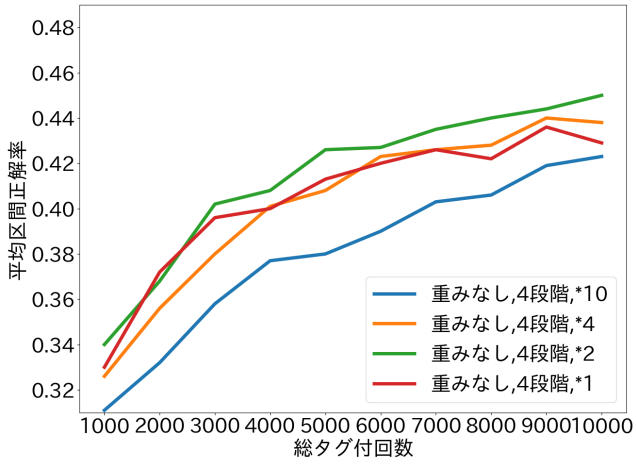




**図 5** 総タグ付け回数を変化させた時の相関係数の推移．*Y は Y 人分のタグが付けられていることを表す．

**Fig. 5** Transition of correlation coefficient when total number of tagging is changed. * Y means that each image is tagged for Y people.



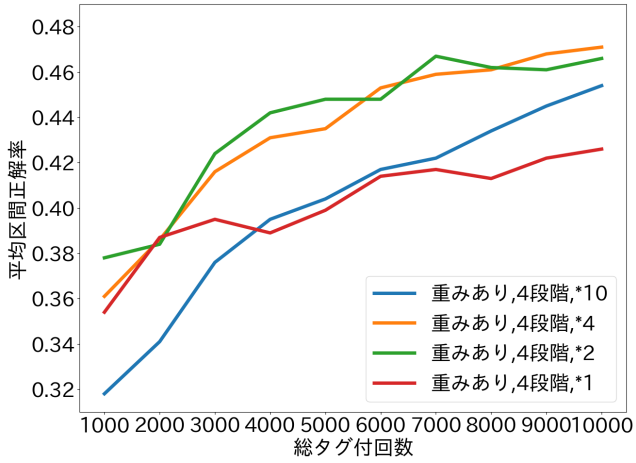




**図 6** 総タグ付け回数を変化させた時の平均区間正解率の推移

**Fig. 6** Transition of the average section accuracy when the total number of tagging is changed.

### 5.2 タグ付けコストと最適なデータの作成方法

#### 5.2.1 目的

総タグ付け回数を変化させたときに，画像枚数と 1 枚にタグ付けする人数の最適なバランスも変化するのか調べる．

#### 5.2.2 手法

総タグ付け回数を 1000 から 10000 まで 1000 ずつ変化させて，1 つ前の実験と同様な学習を行い，結果を比較する．1 つ前の実験で作成した訓練データから必要枚数を抜き出す．抜き出す時は，ランダムに画像を選択して抜き出しており，抜き出し方によって結果が変わってくるので，5 回平均で評価を行う．

#### 5.2.3 結果・考察

**図 5**，**図 6** はそれぞれ 4 段階評価における相関係数，平均区間正解率の推移である．MAE は相関係数と同様の傾向を示したため省略する．また，バイナリ評価も 4 段階評価と同様の傾向を示したため省略する．これらを見ると，総タグ付け回数が少ない時は，画像枚数を優先させた方がよく，総タグ付け回数がある程度増えた段階で 1 枚にタグ付けする人数を増やしたほうがいいことが分かる．

## 6. おわりに

本研究では，FMID を作成し，効果的なデータセットの作成方法を調査した．総タグ付け回数，即ちデータセットの作成にかけられるコストが低い時は，画像枚数を優先したほうがよく，ある程度コストがある時は 1 枚にタグ付けする人数を増やした方が良いことが分かった．また，重み付けを行う際は 1 枚の画像に対して複数人でタグ付けをしていないと逆効果になることも分かった．重み付け以外での，不均衡への対処方法としては，マルチラベルで一気に学習するのではなく，タグごとにダウンサンプリングを行い，半教師あり学習でデータ数を補うといったことが考えられる．

今回行った学習では，1 枚の画像に対して 1 人分のタグしか付けられてないものでも，総勢 10 人で付けたものを用いているので，訓練データの作成に関わる人数を変化させてそれぞれの評価値の推移を調べる必要があると考えている．

© 1959 Information Processing Society of Japan

## 参考文献

[1] Kambe, M., Yokoyama, S., Yamashita, T. and Kawamura, H.: Estimating Impressions for Clothing, Landscape, and Indoor Images Using CNN, Springer, Cham, pp. 67–78 (online), DOI: 10.1007/978-3-030-37442-6_7 (2020).

[2] Yu, A. and Grauman, K.: Fine-Grained Visual Comparisons with Local Learning, *2014 IEEE Conference on Computer Vision and Pattern Recognition*, IEEE, pp. 192–199 (online), DOI: 10.1109/CVPR.2014.32 (2014).

[3] 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, *2015 IEEE International Conference on Computer Vision (ICCV)*, IEEE, pp. 3343–3351 (online), DOI: 10.1109/ICCV.2015.382 (2015).

[4] Liu, Z., Luo, P., Qiu, S., Wang, X. and Tang, X.: DeepFashion: Powering Robust Clothes Recognition and Retrieval with Rich Annotations, *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, IEEE, pp. 1096–1104 (online), DOI: 10.1109/CVPR.2016.124 (2016).

[5] Zheng, S., Yang, F., Kiapour, M. H. and Piramuthu, R.: ModaNet: A Large-scale Street Fashion Dataset with Polygon Annotations, *2018 ACM Multimedia Conference on Multimedia Conference - MM '18*, New York, New York, USA, ACM Press, pp. 1670–1678 (online), DOI: 10.1145/3240508.3240652 (2018).

[6] Vittayakorn, S., Umeda, T., Murasaki, K., Sudo, K., Okatani, T. and Yamaguchi, K.: Automatic Attribute Discovery with Neural Activations, pp. 252–268 (online), DOI: 10.1007/978-3-319-46493-0_16 (2016).

[7] Matsukawa, T. and Suzuki, E.: Person re-identification using CNN features learned from combination of attributes, *2016 23rd International Conference on Pattern Recognition (ICPR)*, IEEE, pp. 2428–2433 (online), DOI: 10.1109/ICPR.2016.7900000 (2016).

[8] Zhao, B., Feng, J., Wu, X. and Yan, S.: Memory-Augmented Attribute Manipulation Networks for Interactive Fashion Search, *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, IEEE, pp. 6156–6164 (online), DOI: 10.1109/CVPR.2017.652 (2017).

[9] Xiong, Y., Li, Y., Pan, P. and Chen, Y.: A regression-based Kansei engineering system based on form feature lines for product form design, *Advances in Mechanical Engineering*, Vol. 8, No. 7, p. 168781401665610 (online), DOI: 10.1177/1687814016656107 (2016).

[10] Vaccaro, K., Shivakumar, S., Ding, Z., Karahalios, K. and Kumar, R.: The Elements of Fashion Style, *Proceedings of the 29th Annual Symposium on User Interface Software and Technology - UIST '16*, New York, New York, USA, ACM Press, pp. 777–785 (online), DOI: 10.1145/2984511.2984573 (2016).

[11] Talebi, H. and Milanfar, P.: NIMA: Neural Image Assessment, *IEEE Transactions on Image Processing*, Vol. 27, No. 8, pp. 3998–4011 (online), DOI: 10.1109/TIP.2018.2831899 (2018).

[12] He, K., Zhang, X., Ren, S. and Sun, J.: Deep Residual Learning for Image Recognition, *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, IEEE, pp. 770–778 (online), DOI: 10.1109/CVPR.2016.90 (2016).

[13] Loshchilov, I. and Hutter, F.: Decoupled Weight Decay Regularization, *International Conference on Learning Representations*, (online), available from ⟨https://openreview.net/forum?id=Bkg6RiCqY7⟩ (2019).

© 1959 Information Processing Society of Japan