

ソーシャルタグを利用した多峰性選好モデル構成のための半教師あり学習

田中 哉汰^{1,a)} 片寄 晴弘^{1,b)}

概要: 本稿は複数の様式、選好基準が存在するデータに対する半教師あり学習を提案する。現在人気を誇るソーシャルメディアにおいて、事業対象として注目され始めている選好データには複数の様式が含まれており、一軸での評価が不向きという曖昧性を内包している。このようなデータを区別せず、選好モデルの構成を試みても各様式に共通な平均的なものしか得ることができない。様々なジャンルで集積されている選好データのうち、本研究ではファッション領域に焦点を当て、相互情報量の最大化による教師なし学習 (IIC) の枠組みを応用した多峰性選好モデル構成の実現に向けた半教師あり学習法を提案する。選好データを予測する回帰問題として評価を実施したところ、提案手法は、ResNet 等、CNN のベースライン的なモデルと比較して、30~40 %の精度向上が認められた。このことから提案手法の有効性とファッション領域が多峰性選好モデルを適用すべき領域対象であることが確認された。

1. はじめに

現在、ソーシャルメディアサービスは絶大な人気を誇っている。その空間上で、テキストや画像、音声や音楽などの多岐にわたるデータが共有されており、「いいね」や「お気に入り」といったソーシャルタグについても日々集積が進んでいる。これらの評価や購入歴を含めたソーシャルタグの利活用は、GAFA 等 IT プラットフォーマーの中心的な事業対象として位置付けられおり、レコメンデーションサービスへの展開も進んでいる [1], [2].

近年、ソーシャルタグ情報は、深層学習における学習用データセットとして利用されるされることも増えてきている [3], [4], [5]. 2017 年の東京ガールズコレクションでは服のスナップショットを撮影して、どの程度「いいね」がつくかを推定する AI 審査員が導入されている [3]. これに先立ち、Simo-Serra らはスナップショットに付与された「いいね」の数をオシャレさと定義し、深層学習を用いた回帰分析の寄与率からオシャレの構成要素を分析している [4].

ソーシャルタグを利用した集合知構成の高精度化に向けては、ソーシャルタグの曖昧さに起因する問題に対処していくことが求められる。一つのジャンルを想定した場合にもそこには (必ずしも意識されとは限らない) 複数の様式が存在し、さらに、それぞれの様式に対しての (必ずしも同一とは限らない) 選好基準が存在する。本研究では、

この問題の解決に向けて相互情報量の最大化による教師なし学習手法 (IIC)[6] の枠組みを応用した、多峰性選好モデル構成のための半教師あり学習法を提案する。ファッション画像を対象とした「いいね」数を予測する課題 (回帰問題) を題材に提案手法の有効性を確認し、多峰性選好問題において様式 (クラス) を取り扱っていく上での課題について議論する。

2. 問題設定と本研究のアプローチ

2.1 選好タグ付与の曖昧性

[3], [4], [5] での応用事例が示すように深層学習におけるソーシャルタグ情報活用には大きな可能性がある。印象や評価にかかるソーシャルタグ情報情報の利用においては、曖昧さや誤差の問題が避けられないが、大規模データとして統計的な特徴を捉えていくことで、全体の傾向の把握や、価値や人気判定システムとしても利用が可能となっている。その一方で、大規模データを用意したとしても解決できない曖昧性が存在する。

選好タグ (いいね) が付与されていく状況を、ファッションジャンルを例にあげて説明する (図 1)。評定者が「いいね」をつける際には、目の前にある評価すべき対象があり、その対象の様式を識別した上で、それぞれの選好基準に基づいて「いいね」タグを与えていく。評定を行う対象に、モード系、ストリート系の様式の素材が含まれているとしよう。前者では「白と黒の対比の綺麗さ」、後者では「緩いコーディネート (の中で組み合わせの良さ)」を選好基

¹ 関西学院大学

^{a)} ddc18531@kwansei.ac.jp

^{b)} katayose@kwansei.ac.jp

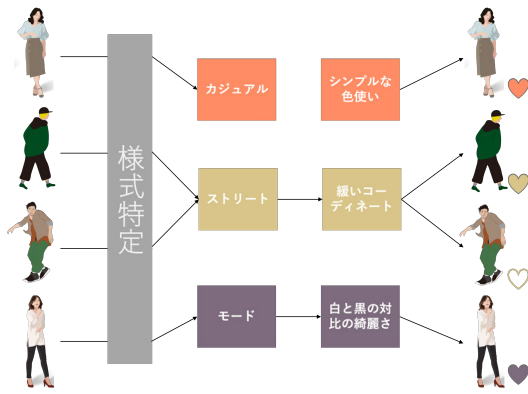


図 1 選好タグの付与の様相.*2
Fig. 1 Preference criteria

準として「いいね」のタグ付けがなされていることが想定される。一般的に、深層学習のデータセットとして用いられる情報は、ソーシャルタグと評定の対象となった素材の組み合わせであり、評定者が想定した「様式」はわからない。その状況で、モード系、ストリート系での例のように選好基準の違いがあると、データセットとしての一貫性は保証されず、選好モデルの構成を試みても各様式に共通な平均的なものしか得られない。この問題は、学習用データセットのサイズを大きくしても解決できるものではない。

2.2 アプローチ

ソーシャルタグを利用した集合知構成の高精度化に向けては、前節に示したような選好タグ付与の様相に合致した学習モデルを用意することが求められる。本論文での基本的なアイデアは、1)「様式」の分類、2)「様式」毎に存在する最適な選好基準のモデル化を構成していく。「様式」の分類については人間の実施内容を模するものを用意できれば理想的であるが、このこと自体きわめて大きな問題である。本研究では、クラスタリング（教師なし学習）として大きなインパクトを与えた Ji [6] らの手法（相互情報量の最大化による教師なし学習手法、以下 IIC）をベースに、クラスタ数の自動推定を含めた「様式」推定機構を提案する。それぞれの様式毎に選好基準推定モデルを配置し、最終段における「いいね」ラベル選好結果が合致するように回帰モデルを構成していく。

2.3 IIC の概要

従来の教師なし学習手法には大きく二つの問題があることが指摘されてきた。1) 特徴量からクラスタリングする際に一つのクラスタに集まってしまう、または、必要なクラスタが消えてしまう。2) 学習データにノイズが含まれ

ている場合に良い特徴量を得られない。

IIC では相互情報量を最大化する手法によって上記問題に対処する。入力データ x のペアとなる x' を x を元に生成し、ニューラルネット ϕ を学習させる。 (x, x') を入力として与え、出力 (z, z') を得て、 $I(z, z')$ を最大化するようにニューラルネットを学習を進める。(式 1)

$$\text{Max } I(z, z') = H(z) - H(z|z') \quad (1)$$

$H(z)$ は z に対するエントロピーを表しており、 z が一様分布に近づくように機能するため 1) の特定のクラスタに集まってしまうことに対処できる。 $H(z|z')$ は z' が与えられたときの条件付きエントロピーを表しており、出力に意味合いを持たせるように学習が進むため 2) に対して有効に機能する。相互情報量を用いて従来の教師なし学習において課題とされていた部分を解決することで、既存のモデルと比較して高い精度を記録している。

また、画像分類のベンチマーク (STL-10)[7] 実験から「半教師あり学習に IIC を適用することで、教師あり学習の精度を超える」可能性が示唆されている。多峰性モデルを取り扱っていく上では極めて有望な特質であるが、ソーシャルタグに対しての選好モデル構成に向けての応用研究は現時点で他に存在していない。

3. 複数様式に対する多峰性選好モデルの構成

図 2 に提案モデルを示す。IIC をベースにクラスタ数自動推定を含めた「様式」推定機構を構築する。提案手法は二段階のカスケードモデルであり、前段階では IIC を応用したクラスタリング、後半では様式ごとの違いを考慮した選好基準推定を行う。スナップショットの入力データに対して「いいね」の予測選好ラベルが合致するように最適化を図るモデルを提案する。

初めに、1) IIC の手法に則り入力データからクラスタ数が $k = 3, \dots, 12$ の確率分布をそれぞれ生成する。次に、2) 生成された確率分布から最適なクラスタ数 k を決定する (IIC を応用したクラスタリング)。最後に、3) CNN によって入力データの回帰分析をすると共に、生成された確率分布から OneHot ベクトルを作成し、入力データを学習させている CNN における線形層の新たな特徴量として OneHot ベクトルを結合することで、選好基準推定を行う。

1) IIC では元画像 x から変換を加えたものを $x' = g(x)$ としており、幾何学的、色彩的な変換を加えている。幾何学的な変換では、スケーリング、スキュー、回転、反転の 4 種類、色彩的な変換ではコントラスト変換、彩度変換の 2 種類、合計 6 種類の変換をランダムに加えている。本研究でも上記のスキームに則り、データセット $X = x_1, x_2, \dots, x_n$ の画像 x からランダムに変換を加えた x' を生成する。(図 3)

2) IIC では、クラスタ数 k を予め決定する必要があ

*2 著作権および肖像権保護のために著作権フリーのイラストを代用している。

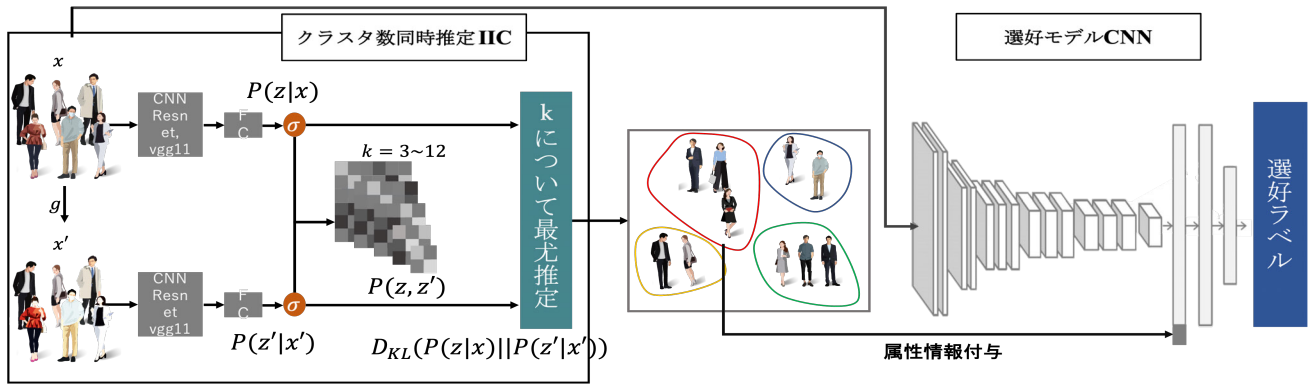


図 2 IIC を応用したクラスタ数同時推定部分と選好モデル部分から成るカスケードモデル. *2

Fig. 2 Cascade model consisting of a simultaneous estimation of the number of clusters part using IIC and a Preference model part.



図 3 変換事例. *2

Fig. 3 Image change(g) example

る。本研究においては、複数の様式を含むデータセットを様式ごとのクラスタに分解したいが、様式の数是不定である。IIC は相互情報量を最大化するように学習を進めるため、生成される確率分布 $P(z|x)$ と $P(z'|x')$ は徐々に近いものになることが明らかになっており、本研究ではこの性質を利用したクラスタ数の決定を行う。当てはまりとモデルの自由度、双方のバランスをとることを考慮し、確率分布 $P(z|x)$ と $P(z'|x')$ が最も近い状態になるように KL-divergence を用いることで k を定める。(式 2)

$$\text{Min } D_{KL}(P(z|x)||P(z'|x')) \quad (2)$$

3) 多峰性選好モデルにおいて「様式」が観測できることは重要なトピックであり、本研究においても焦点を当てている。しかし、従来の CNN モデルではこういった「様式」情報はブラックボックス化しており、様式 (クラスタ) がどのようにコーディングされているのかを確認することができない。

画像認識系の学習済みモデルにおける Layer の分析研究によって「色情報」あるいは「テキスト情報」等、そ

の layer が担っている特徴が何かわかりつつある [8], [9]. Robert らは ImageNet という 1400 万枚以上のデータセットで学習させた畳み込みニューラルネットワークモデルは物体の形状ではなくテキスト情報において分類していることを示しており [8], 畳み込み層において様式を特徴として表現できることが示唆されている。一方で、回帰分析においては畳み込み層において点数に結びつく特徴を捉えるように学習が進行するため、中間層で様式を表現することは期待できない。

そこで提案手法では、CNN を用いた回帰分析に加えて、2) において生成した 確率分布を OneHot ベクトルに変換し、CNN モデルの線形層に新たな特徴量として結合する。OneHot ベクトルを特徴量として追加することで学習に指向性を持たせられることが既に明らかになっており、様々な研究で活用されている [10], [11]. これによって様式 (クラスタ) を考慮したまま学習を進めることで、多峰性選好モデルの構築が可能となる。

4. 実験

4.1 実験設定

ファッション画像を対象とした「いいね」数を予測する回帰タスクにおいて有効性の検証を行う。それに先立ち、交差検証、アブレーション分析を実施する。

本研究では、複数の様式とその様式ごとの選好基準が存在し、それぞれの基準に対して選好されたデータ (いいねフラグのオン・オフが付けられた) データセットが観測され、かつ、様式情報については観測されないという条件設定における手法を提案している。ソーシャルメディア上のスナップショットに対して、不特定多数の人が「いいね」の評価をしており、その数が観測されている。このようなデータは先述している条件設定、複数の様式とその様式ごとの選好基準が存在し、それぞれの基準に対して選好されたデータに該当する。

条件設定を満たすデータセットを作成し、様々なベ

ンチマークにおいて安定して高い精度を記録している5つのモデル [Alexnet[12], Densenet[13], Inception v3[14], ResNet18[15], vgg11[16]] にて比較する。

提案手法の有効性検証のため、提案手法と図2においてクラスタ推定および属性付与レイヤを除いたもの（以下、基本モデル）との比較実験を実施する。提案手法における選好モデルはCNNをベースに構成されており、基本モデルとの比較のために同一のモデルに適宜変更し評価を行う。評価指標には二乗平均平方根誤差 (RMSE), 平均絶対誤差 (MAE) を使用する。

また、汎化性能を評価するため提案手法において、交差検証を行う。データ群を c 個に分割し、そのうち1つをテストデータ、残り $c-1$ 個を学習データとし、精度検証を c 回繰り返す。

4.2 データセット

様式を考慮したモデル構成の検証を行うためにデータセットを作成する。データセットは複数の様式が混ざっている、かつ、選好ラベルが付与されている必要があり、これらの要件を満たすデータをユーザ投稿型のファッションソーシャルサイト Wear から収集する。このサイトでは、ユーザは自分自身のスナップショットを投稿しており、それぞれの写真は異なる角度からユーザの姿が映し出されていたり、顔がわからないようなボカシが入っていたりする。他のユーザはこれらの投稿を見て、コメントの投稿や「いいね」、「お気に入り」といったタグ付けを行うことができ、また、ユーザのフォロワーになることができる。

ユーザの投稿に対して制限がないため、人が写っていないものや複数人数で撮影している投稿が存在しており、これらの画像を削除することで43,540件のデータセットを作成する。本研究ではスナップショットとその画像に付随した「いいね」の数を選好データとして使用する。

学習の安定性や、評価の公平性を担保するためにはデータの外れ値に対処する必要がある。収集したデータは投稿時期やフォロワー数の多寡によって、いいね数が大きく外れた値が存在するため、データの正規化を行う。

Simo-Serraらはユーザ投稿型のソーシャルサイトからデータセットを作成する際に、選好データの偏りに対して、月ごとのヒストグラムを生成して正規分布に変換することでデータの偏りに対処した。[4] 本研究でも Simo-Serraらと同様に月ごとのヒストグラムに対して、box-cox変換(式3)をおこなった後、min-max正規化(式4)をおこなうことで0~100のスケールに納まるように変換する。

$$f(x)^\lambda = \begin{cases} \frac{x^\lambda - 1}{\lambda} & (\lambda \neq 0) \\ \ln(x) & (\lambda = 0) \end{cases} \quad (3)$$

表1 精度比較
Table 1 Accuracy

モデル名	評価指標	提案手法	基本モデル
Alexnet[12]	RMSE	12.53	18.40
Densenet[13]		12.20	17.77
Inception v3[14]		14.69	18.26
ResNet18[15]		11.51	17.87
vgg11[16]		12.82	17.83
Alexnet[12]	MAE	10.43	15.88
Densenet[13]		9.35	16.24
Inception v3[14]		10.83	17.10
ResNet18[15]		8.72	14.65
vgg11[16]		9.10	16.30

$$x = \frac{good - good_{min}}{good_{max} - good_{min}} \quad (4)$$

4.3 実験結果

4.1章、4.2章で述べた実験環境で得られた結果を以下の表1に示す。縦軸が比較に用いたモデル名、横軸では評価指標と、属性情報を付与した提案手法と基本モデルの比較を示している。表1における数値は交差検証における平均値を表している。

全てのモデルにおいて、提案手法の方が基本モデルよりも良い性能であることが確認できた。提案手法における平均精度はRMSEにおいて12.75、MAEでは9.68であった。どちらの評価指標においても最も精度が良かったモデルはResNet18[15]であり、RMSEにおいては35.6%、MAEにおいては40.6%の精度向上しており、また、RMSEにおいては平均で29.24%、MAEでは平均39.64%の精度向上が確認できた。

交差検証による評価結果を表2に示す。RMSE、MAEにおける交差検証の結果を表しており、縦軸が $c=5$ 交差検証による分割の結果、横軸は評価したCNNモデルを示している。どちらの評価指標においても大きく外れた結果が存在せず、データセットをどのように分割しても安定した学習を行えていることが窺える。

モデルは最も高い精度を記録していたResNetを使用し、交差検証における平均結果を表している。幾何学変換、色彩変換の両方を加えたモデルが最も精度が高く、色彩変換だけを加えたモデルの方が、幾何学変換のみを加えたモデルよりも高精度であった。

今回使用したデータセットに対して提案モデル前段により求められたクラスタ(様式)数は6であった。(クラスタリングの結果については著作権および肖像権保護のために割愛している。)特定のクラスタにおいては顕著な特徴が現れており、あるクラスタでは他のクラスタと比較して明るい色を用いたファッションが多い。また、特定のクラスタにおいては柄を用いたものが多く、一方で、2つのクラスタにおいてはシンプルなファッションが多く分類された。

表 2 交差検証
Table 2 Cross Validation

	評価指標	Alexnet	Densenet	Inception	ResNet	vgg
c=1	RMSE	12.29	12.43	13.88	12.42	13.74
c=2		11.22	12.29	14.51	11.25	12.50
c=3		12.84	12.83	14.19	11.39	14.34
c=4		14.06	11.51	16.32	10.96	11.77
c=5		12.04	11.94	14.55	11.53	11.75
average		12.53	12.20	14.69	11.51	12.82
c=1	MAE	9.62	9.15	11.08	7.76	8.65
c=2		10.07	10.84	10.29	9.10	9.91
c=3		11.38	7.99	10.59	7.91	7.65
c=4		11.48	9.05	12.66	10.41	10.57
c=5		9.60	9.72	9.53	8.42	8.72
average		10.43	9.35	10.83	8.72	9.10

4.4 実験結果の総括

実際に人々が「いいね」を付けている対象を考えれば、同一ジャンルであっても複数の様式が存在していることが通常である。ファッション領域で言えば、ストリート系やゴスロリ等の異なったスタイルが存在し、これらスタイル毎に異なった選好基準（価値観）が存在している。「いいね」は、選好基準（価値観）を満たすものに与えられるが、その選好基準でのスタイルがなんであったについては直接観測できないことが少なくない。その対象共通の選好基準もあろうが、様式毎に反駁する基準もあり得る。この前提で、学習を試みても十分なパフォーマンスが得られない可能性がある。本研究では、4.2節で議論した条件設定において IIC の枠組みを応用したスタイル別選好モデルの構成手法は上記の問題解決手段として有効であるという仮説を立て、評価をおこなった。

「いいね」の数を予測する回帰タスクの形で提案手法の検証を行った結果、表 1 が示すように、MAE における平均精度が 9.68 であり、これは平均 10 点未満の誤差でスナップショットに対して性能評価できることを示している。MAE よりも RMSE の方が数値が全体的に大きいことから、予測全体において分散が大きいことが読み取れ、選好データに対する回帰分析の難しさが表れている。また、既存のモデルと比較して 30~40% 程度高い精度で分類できており、様式毎の属性情報を考慮したモデル構築はパフォーマンスを伸ばすのに有効であることが確認できた。

また、様式のクラスタリングにおいては、特定のクラスにおいて一定の傾向があることが確認できた。これは提案手法における多方向性様式の同時推定が人が捉えている様式の分類を可能にしていることを示唆している。一方で $k=1$, $k=2$ のクラスタに含まれるスナップショットはコンサバ系のファッションに該当し、類似しているように見える。これは IIC の性質上の類似画像が別々のクラスタに含まれているものだと考えられる。IIC では既存のクラスタ分類において特定のクラスタに集まりすぎてしまうといった問題を一様分布に近づけるように学習させることで対処してい

る。つまり、全てのクラスタに分類される画像数は均等になるように学習されるため、 $k=1$, $k=2$ に見られるようなファッションは人気のファッションであり、データセットの中でも多くの数を占めているため、別々のクラスタに分布されたものと推察される。

5. 議論

5.1 提案手法について

ファッション領域において提案手法は 30~40% の性能向上が見られたことから、この領域が本来、多峰性選好モデルを適用すべき領域対象であったことが確かめられた。また、同時に提案手法の有効性も確認できた。

本研究においては多峰性選好データの一例としてファッションに焦点を当てている。様々な対象に対してソーシャルタグの集積は進んでおり、また、ソーシャルタグの背景となっているものは人の感性（選好モデル）は本質的に多峰性を考慮すべきテーマであるため応用範囲は広いと考えられる。加えて、提案手法は単純に好きか嫌いかというデータに対しても有効に働く。

本研究においては、データセット内に存在する潜在的な「様式」への着眼がポイントであり、相互情報量の最大化によって得られる情報論的な基準で設定されるクラスタを「様式」と仮定してモデル化を実施した。これは、人間が捉えている「様式」とは必ずしも同じではなく、これを捉えていくことが今後の課題である。

5.2 今後の展望

カスケードモデル前段のみによる最適化と、系統全体での最適化によって求められている様式（クラスタ）数が異なることが明らかになり、精度向上には選好データを最大限活かすために、様式と選好モデルの同時推定を行っていく必要がある。そこで、カスケードモデルの前段、IIC を応用したクラスタリングをノンパラメトリックベイズのアプローチによって定式化することを目指す。これによって、一様分布に寄りすぎてしまうために、似たような「様式」を別々のクラスタに分けてしまうといった問題を解決する方策の一つになることが期待できる。

本研究のアプローチは選好モデルを並列に繋いだものではなく、OneHot ベクトルによって指向性を持たせるようなアプローチであり、複数のモデルを構築しているわけではないためノンパラメトリックベイズのアプローチを応用させやすい。そのため、クラスタリング部分に留まらず、最終的な展望は、ノンパラメトリックベイズに基づいてカスケードモデルの様式（クラスタ）数と嗜好モデルを同時に推定するモデルの構築に取り組むことである。

6. おわりに

本研究では複数の様式、選好基準が存在するデータにお

ける半教師あり学習を提案した。絶大な人気を誇るソーシャルメディアにおいて、現在注目され始めている選好データには複数の様式が含まれていることに着目し、当てはまりとモデルの自由度、双方のバランスを取ることでIICを改良し、上記課題に対処した。

提案手法において、何件の「いいね」が期待できるかを予測する回帰問題として評価を実施したところ、基本モデルと比較して、概ね40%の精度向上が見られた。この結果から複数の様式と選好基準を含むデータセットに対して、提案手法がパフォーマンスを向上させる上で有効であることが確かめられた。また、ファッション領域は多峰性選好モデルを適用すべき領域対象であったことも確かめられた。

情報論的な基準で設定した「様式」と人間が捉えている「様式」は必ずしも同じでない。ノンパラメトリックベイズを応用することで、様式と多峰性選好モデルの同時推定モデルを構築することが今後の課題である。

参考文献

- [1] Jianmo Ni, Jiacheng Li, and Julian McAuley. Justifying recommendations using distantly-labeled reviews and fine-grained aspects. In *(EMNLP-IJCNLP)*, pages 188–197, Hong Kong, China, November 2019. Association for Computational Linguistics.
- [2] Paul Covington, Jay Adams, and Emre Sargin. Deep neural networks for youtube recommendations. In *Proceedings of the 10th ACM Conference on Recommender Systems, RecSys '16*, pages 191–198, New York, NY, USA, 2016. Association for Computing Machinery.
- [3] 株式会社マイナビ, TOKYO GIRLS COLLECTION SNAP CONTEST, (最終閲覧日: 2021年10月4日). <https://tgc.girlswalker.com/17aw/special/snap/>.
- [4] Edgar Simo-Serra et al. Neuroaesthetics in fashion: Modeling the perception of fashionability. In *2019 10th (IEEE)*, 2016.
- [5] 森 鈴果, 田中 哉汰, 橋田 光代, and 片寄 晴弘. 恋ダンスの達人: 深層学習を利用したダンスのキレ採点システム. In *研究報告エンタテインメントコンピューティング*, volume EC-59(2), 2021.
- [6] Xu Ji, Andrea Vedaldi, and Joao Henriques. Invariant information clustering for unsupervised image classification and segmentation. In *2019 (ICCV)*, pages 9864–9873, 2019.
- [7] Adam Coates, Andrew Ng, and Honglak Lee. An analysis of single-layer networks in unsupervised feature learning. In Geoffrey Gordon, David Dunson, and Miroslav Dudík, editors, *Proceedings of the Fourteenth International Conference on Artificial Intelligence and Statistics*, volume 15 of *(PMLR)*, pages 215–223, Fort Lauderdale, FL, USA, 11–13 Apr 2011.
- [8] Robert Geirhos et al. Imagenet-trained cnns are biased towards texture; increasing shape bias improves accuracy and robustness. volume abs/1811.12231, 2018.
- [9] Gatys et al. Image style transfer using convolutional neural networks. In *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 2414–2423, 2016.
- [10] Hartatik, Hanif Al Fatta, and Utsman Fajar. Captioning image using convolutional neural network (cnn) and long-short term memory (lstm). In *2019 (ISRITI)*, pages 263–268, 2019.
- [11] Duc-Hong Pham and Anh-Cuong Le. Exploiting multiple word embeddings and one-hot character vectors for aspect-based sentiment analysis. *International Journal of Approximate Reasoning*, 103:1–10, 2018.
- [12] Alex Krizhevsky et al. Imagenet classification with deep convolutional neural networks. In F. Pereira, C. J. C. Burges, L. Bottou, and K. Q. Weinberger, editors, *Advances in Neural Information Processing Systems 25*, pages 1097–1105. Curran Associates, Inc., 2012.
- [13] G. Huang, Z. Liu, L. Van Der Maaten, and K. Q. Weinberger. Densely connected convolutional networks. In *2017 (CVPR)*, pages 2261–2269, 2017.
- [14] Christian Szegedy, Vincent Vanhoucke, Sergey Ioffe, Jonathon Shlens, and Zbigniew Wojna. Rethinking the inception architecture for computer vision. In *CVPR*, 2016.
- [15] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. *arXiv preprint arXiv:1512.03385*, 2015.
- [16] Karen Simonyan and Andrew Zisserman. Very deep convolutional networks for large-scale image recognition. volume abs/1409.1556, 2014.
- [17] 寸田 菜月, 谷 伊織, 飛谷 謙介, 竹本 敦, and 谿 雄祐. Neural style feature を用いた感性モデルに基づく質感表現. In *信学技法*, volume MVE2020 9-20, 2020.