

顔特徴を用いたポートレート写真に対する審美的品質推定

Aesthetic Quality Estimation using Face Features for Portrait-Image

泉水 長門[†] 滝本 裕則^{††} 山内 仁^{††} 金川 明弘^{††} 遠部 雅弘^{†††}

Nagato Sensui[†] Hironori Takimoto^{††} Hitoshi Yamauchi^{††} Akihiro Kanagawa^{††} Masahiro Ombe^{†††}

[†] 岡山県立大学大学院 情報系工学研究科 ^{††} 岡山県立大学 情報工学部 ^{†††} 有限会社バサラ

1 はじめに

ソーシャルメディアにおける写真の大部分は人の顔を含んでおり、私たちが人の顔を含む写真を目にする機会は増加している。また、人が主観的に良いと評価するような写真を撮影したいというニーズは今後ますます高まっていくと考えられる。したがって、人の顔を含む写真の審美的品質評価が求められている。

写真の審美的品質評価とは、人が感覚的に写真を好むかどうかを定量化したものである。これまでに、写真の審美的品質の自動評価を目的として、様々な手法が提案されている[1]-[4]。近年では、画像認識分野において高い性能を示しているCNN(Convolutional Neural Network)を用いた手法[5]-[9]が提案され、その有効性が示されている。

これらの既存研究は、写真中に人の顔を含むかどうかは関係なく、一般的な写真全てを対象としている。しかし、人の顔を含む写真について人が審美的品質評価を行う場合、顔の表情や写真における顔の位置などの顔特徴も考慮して写真の評価を行うと考えられる。したがって、顔を含む写真を対象とした審美的品質の自動評価を行う場合、顔に関係する特徴量を用いることにより審美的品質の推定精度が向上すると考えられる。

本稿では、人の顔を含む写真(以下、ポートレート写真という)について、より精度の高い審美的品質評価を実現するために、顔に関する特徴とCNNより抽出される画像全体の特徴(以下、グローバル特徴という)を用いて審美的品質を推定する手法を提案する。また、顔に関係する特徴は多様であり、抽出した特徴の中には審美的品質を評価するための説明変数として不十分な特徴もあると考えられる。したがって、審美的品質の予測モデルを最適化するため、審美的品質推定に寄与する特徴を特定する特徴選択を行う。

2 提案手法

2.1 顔特徴を用いた審美的品質推定手法

提案手法では、ポートレート写真について、顔特徴およびCNNを用いて審美的品質に関するグローバル特徴を抽出する。

顔特徴の抽出手法として、Microsoft社の提供するFace API[10]、Google社の提供するCloud Vision API[11]を用いる。これらのサービスは、写真に対して顔検出を行い、写真中の顔の位置情報や角度、顔のぼけ具合、感情などの顔特徴の予測値を返す。提

案手法では、Face APIとCloud Vision APIにより抽出される特徴のうち、40種類の顔特徴を用いる。提案手法で用いる顔特徴については、評価実験の項で説明する。

審美的品質に関するグローバル特徴の抽出には、Omoriら[9]の提案したマルチタスクXceptionを用いる。Xception[12]は、Inception[13]-[15]の派生モデルであり、画像認識タスクにおいて高い性能を報告している。Omoriら[9]は、審美的品質推定における審美スコアの平均と投票分布の標準偏差には関連性があると考え、Xceptionを拡張し、審美スコアの平均値と標準偏差を同時に推定するモデルを提案しており、推定精度の向上を報告している。OmoriらのマルチタスクXceptionの概略図を図1に示す。具体的には、Xceptionの全結合層を、写真の審美スコアの平均値の推定と標準偏差の推定の2つに分岐させたモデルである。畳み込み層であるXceptionでは両方のタスクで重みを共有しており、分岐後の全結合層では各タスクで独自に学習される。Xceptionの重みの初期化にはImageNet[16]の重みを用いてグローバル特徴を抽出する。マルチタスクXceptionの審美スコア推定部分において、全結合層の256次元の特徴量を抽出し、審美的品質に関するグローバル特徴として用いる。

提案モデルの概略図を図2に示す。本研究では、データセットにおける審美スコアの平均点を推定するために、サポートベクター回帰(SVR)を用いる。顔特徴とグローバル特徴をSVRの入力とし、審美スコアの推定を行う。 N_f を特徴の総数、 X_j ($j \in [0, N_f]$)を特徴量、 $\phi(x)$ を非線形写像、 w_j を重み、 b をバイアス項とすると、SVRによる審美的品質の予測値 $pred$ は次のように表現される。

$$pred = \sum_{j=0}^{N_f} \phi(X_j) w_j + b \quad j \in [0, N_f] \quad (1)$$

また、重み w はLagrange乗数 $\alpha, \hat{\alpha}$ を用いて次のように表現される。

$$w = \sum_{i=0}^{N_f} (\alpha_i - \hat{\alpha}_i) \phi(X_i)^T \quad (2)$$

2.2 特徴選択手法

顔に関係する特徴は多種多様であり、抽出した特徴の中には審美的品質を評価するための変数として不十分な特徴もあると考えられる。そこで、ランダムフォレストを用いて特徴選択を行うことで予測モデルの最適化を試みる。ランダムフォレストは、決定木を用い

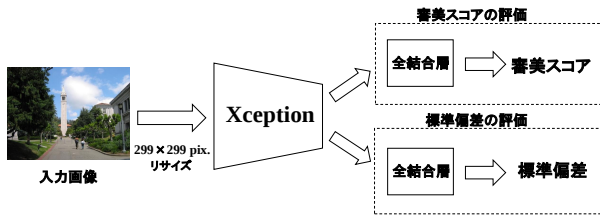


図 1: Omori らのマルチタスク Xception による審美的品質推定モデル [9]

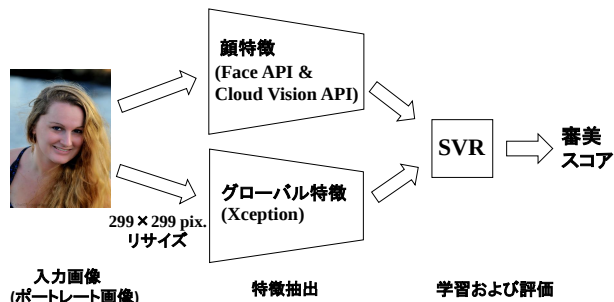


図 2: 提案モデルの概略図

たアンサンブル法の 1 つである。決定木を基に作成されるランダムフォレストでは、各特徴の特徴重要度の平均値を算出することができる。ランダムフォレストにより算出される各特徴の重要度を用いて特徴選択を行うことで、予測モデルの最適化を試みる。具体的には、審美的品質の推定を行う際に、特徴重要度の高いものから順に用いる。特徴選択を行う場合の提案モデルの概略図を図 3 に示す。

3 評価実験

3.1 実験条件

評価実験には、AVA-dataset[17] を用いた。AVA-dataset の各写真には約 200 人の異なる人物により投票された 10 段階の審美スコア (1~10) のヒストグラムが与えられている。ここで、各写真のスコアごとの投票数を $V = [v_1, v_2, \dots, v_{10}]$ とする。 v_i はスコア i の

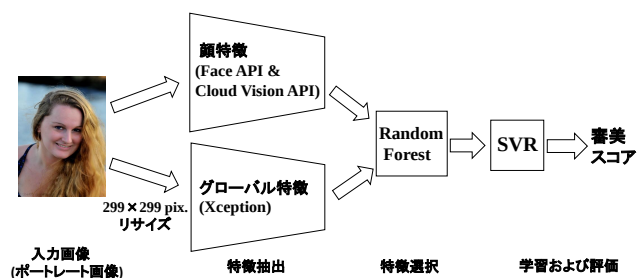


図 3: 特徴選択を行う場合の提案モデルの概略図

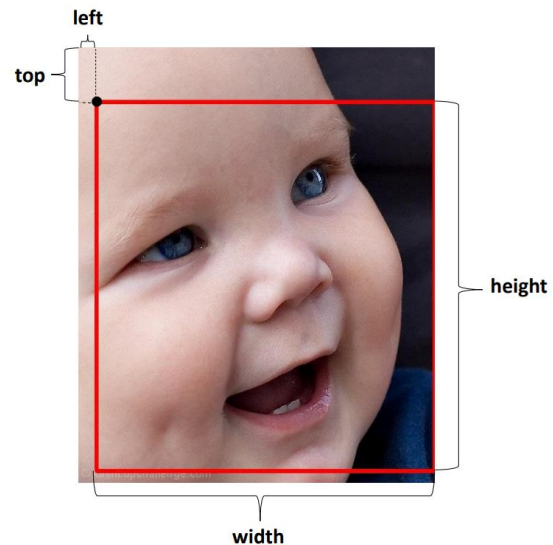


図 4: 顔領域特徴の表現方法

投票数を示す。各写真の審美度 $Score$ と標準偏差 SD は以下のようにして与える。

$$Score = (\sum_{i=1}^{10} i * v_i) / \sum_{i=1}^{10} v_i \quad (3)$$

$$SD = \sqrt{(\sum_{i=1}^{10} (i^2 - Score^2) * v_i) / \sum_{i=1}^{10} v_i} \quad (4)$$

また、各写真にはタグが付けられている。本研究では、portrait タグの写真のうち、顔検出および顔特徴の抽出を行うことができた写真を用いて評価実験を行う。評価実験に用いる portrait タグの写真の枚数は 7,134 枚となった。portrait タグの写真を Face API と Cloud Vision API にそれぞれ入力し、顔特徴を抽出する。ここで、Face API, Cloud Vision API により得られた顔特徴は、数値ではないものが含まれるため、それらの数値化を行う。また、全ての特徴について 0~1 の範囲に正規化を行う。具体的な顔特徴の数値化および正規化方法について、表 2 に示す。顔の位置情報については、顔検出精度の特徴量を出力する Cloud Vision API の値を用い、図 4 のように顔領域の表現方法を変更する。顔領域の端点のうち、画像の原点と原点から最も近い端点との x 方向と y 方向の距離をそれぞれ top と left のパラメータで表現する。また、顔領域の幅と高さをそれぞれ width と height のパラメータで表現する。

portrait のタグ以外の写真は、グローバル特徴抽出のためのマルチタスク Xception の転移学習に用いる。しかし、AVA-dataset に含まれる写真の平均スコアには偏りがある。そのため、平均スコアごとの分布枚数になるべく均一になるように前処理を行う。具体的には、AVA-dataset の写真を審美スコア 0.1 ごとに分類し、あるスコアカテゴリの写真が 1,000 枚を超える場合は 1,000 枚になるようにランダムに選り出し、1,000 枚に満たない場合は反転画像を加えることで画像枚数を増加させた。前処理によって、マルチタスク

Xception の転移学習に用いる写真の枚数は 45,712 枚となった．マルチタスク Xception の学習の際，入力画像サイズは 299×299 画素，学習のエポック数は 100，バッチサイズは 19，Dropout 率は 0.5，最適化手法として Adagrad を用いた．学習したマルチタスク Xception を用いて，portrait タグの写真から 256 次元のグローバル特徴を抽出する．

審美的品質の学習および評価を行う SVR について，カーネル関数は RBF カーネルを用いる．

$$K_{RBF} = \exp(-\gamma \|X_i - X_j\|^2) \quad (5)$$

ここで，SVR のハイパーパラメータである C と γ はグリッドサーチ法による最適化により $C = 10, \gamma = 10^{-3}$ とした．

特徴選択で用いるランダムフォレストについて，木の数を多く設定しすぎると計算に必要なメモリの量も訓練にかかる時間も増大するため，評価実験では木の数を 100 とした．木の最大深さは，全てのノードが 1 つの回帰値しか含まない状態になるまでとした．特徴量の最大数は，用いる特徴の総数 N_f とした．

推定精度の評価指標には，MAE(Mean Absolute Error) と LCC(Linear Correlation Coefficient) を用いる．

3.2 結果と考察

提案モデルの性能評価を行う．比較法として Omori らのマルチタスク Xception[9]，マルチタスク Xception によって抽出した 256 次元特徴のみを用いて SVR による推定を行った結果を用いる．各モデルの推定結果を表 1 に示す．表 1 より，グローバル特徴のみを用いて推定を行うモデルと比較して，顔特徴を用いた提案モデルの方が推定精度が高いことを確認した．以上の結果から，ポートレート写真に対する審美的品質推定において，顔特徴を用いることの有効性を示した．

また，特徴選択を行うことにより提案モデルの推定精度がわずかに向上することを確認した．図 5 に提案モデル（特徴選択）における MAE の推移を示す．横軸が推定に用いた特徴量の数，縦軸が MAE である．図 3 より，特徴量数が 180 の付近で推定精度が収束していることが分かる．表 1 より，特徴量を大幅に削減した場合でも，Xception + SVR モデルの推定性能を上回っていることを確認した．また，特徴選択を行わない提案モデルと比較して，推定精度はほとんど変化しないことを確認した．以上のことから，特徴選択は有効であったといえる．

ランダムフォレストの特徴選択によって算出された特徴重要度の値が比較的高いものとして，顔の領域情報 (fdBoundingPoly)，顔検出精度 (detectionConfidence)，ピッチ角以外の角度情報，感情のはっきりしない (neutral)，顔のぼけ具合 (blur)，顔の露出具合 (exposure)，ノイズの具合 (noise) が挙げられる．これらの特徴量は，ポートレート写真に対する審美的品質推定において，特に有効な特徴量であると考えられる．

表 1: 各モデルの推定結果

条件		評価指標	
モデル	特徴数	MAE	LCC
Xception[9]		0.57729	0.55309
Xception[9] + SVR	256	0.45332	0.55967
提案モデル	296	0.45140	0.56457
提案モデル (特徴選択)	293	0.45138	0.56459
提案モデル (特徴選択)	180	0.45207	0.56327

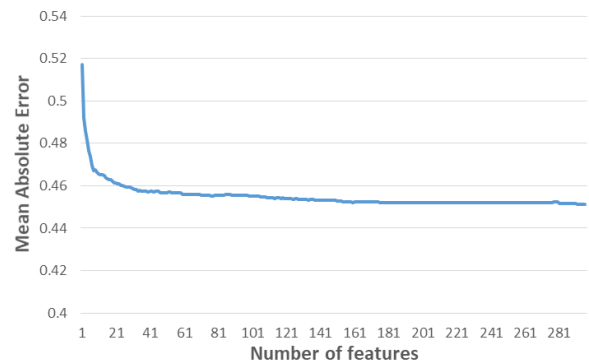


図 5: 提案モデル（特徴選択）における MAE の推移

図 6 に本研究において，最も推定性能の高かった提案モデル（特徴次元数 293）によるポートレート画像の審美スコアの予測値の例を示す．

4 おわりに

本稿では，ポートレート写真に対する審美的品質推定の高精度化を目的とし，顔特徴および CNN によって抽出されるグローバル特徴を用いて審美的品質の推定を行う手法を提案した．Microsoft Azure Face API と Google Cloud Vision API を用いて抽出した顔特徴およびマルチタスク Xception によって抽出されるグローバル特徴を用いて，SVR によって審美的品質推定を行った．また，審美的推定の予測モデル最適化を行うために，ランダムフォレストを用いた特徴選択の手法を提案した．

評価実験の結果，顔特徴を考慮した提案モデルによる推定は，CNN より抽出されるグローバル特徴のみを用いたモデルによる推定よりもより高い性能であることを示した．また，ランダムフォレストによる特徴選択を行った提案モデルは，特徴選択を行わない場合よりもわずかに高い性能であることを示した．

なお，本研究の一部は，JSPS 科研費 18K11402 の助成を受けて行われたものである．

表 2: 使用した顔特徴および正規化方法

	特徴量	意味	正規化前の値	正規化後の値
Microsoft Face API[10]	blur(value)	顔のぼけ具合	0～1	
	anger	怒り	0～1	
	contempt	軽蔑	0～1	
	disgust	嫌悪	0～1	
	fear	恐怖	0～1	
	hapiness	幸せ	0～1	
	neutral	感情のはっきりしない	0～1	
	sadness	悲しみ	0～1	
	surprise	驚き	0～1	
	exposure(value)	顔の露出具合	0～1	
	eyeMakeup	目のメイクの有無	0～1	
	lipMakeup	唇のメイクの有無	0～1	
	noise(value)	ノイズの具合	0～1	
	beard	顎ひげ	0～1	
	moustache	口ひげ	0～1	
	sideburns	もみあげ	0～1	
	bald	禿げているかどうか	0～1	
	invisible	髪が見えているかどうか	0～1	
	pitch	ピッチ角（顔角度）	0	pitch / 90
	roll	ロール角（顔角度）	-45～45	(roll + 45) / 90
	yaw	ヨー角（顔角度）	-45～45	(yaw + 45) / 90
	eyeOccluded	目の塞がれ具合	0～1	
	foreheadOccluded	額の塞がれ具合	0～1	
	mouthOccluded	口の塞がれ具合	0～1	
	smile	笑顔であるかどうか	0～1	
Google Cloud Vision API[11]	fdBoundingPoly	皮膚輪郭の矩形領域	矩形の頂点 4 点の座標	図 4 のように変更後、 画像サイズで 0～1 の範囲に正規化
	detectionConfidence	顔検出精度	0～1	
	rollAngle	ロール角	-45～45	(rollAngle + 45) / 90
	panAngle	パン角	-90～90	(panAngle + 90) / 180
	tiltAngle	ティルト角	-45～45	(tiltAngle + 45) / 90
	joyLikelihood	楽しみの度合い	VELY_UNLIKELY	VELY_UNLIKELY = 0
	sorrowLikelihood	悲しみの度合い	UNLIKELY	UNLIKELY = 0.25
	angerLikelihood	怒りの度合い	POSSIBLE	POSSIBLE = 0.5
	surpriseLikelihood	驚きの度合い	LIKELY	LIKELY = 0.75
	underExposedLikelihood	肌の露出の度合い	VERY_LIKELY	VELY_LIKELY = 1
	blurredLikelihood	ぼけの度合い		
	headwearLikelihood	何かを被っている度合い	の 5 値	に変換

参考文献

- [1] Y. Luo, and X. Tang: "Photo and Video Quality Evaluation: Focusing on the Subject", Proc. of European Conference on Computer Vision, pp. 386-399 (2008)
- [2] M. Nishiyama, T. Okabe, I. sato, and Y. Sato: "Aesthetic Quality Classification of Photographs Based on Color Harmony", Proc. of IEEE Computer Society Conference on Computer Vision and Pattern Recognition, pp. 33-40 (2011)
- [3] D. Joshi, R. Datta, E. Fedorovskaya, Q. T. Luong, J. Wang, J. Li, and J. Luo: "Aesthetics and Emotions in Images", IEEE Signal Processing Magazine, Vol. 28, No. 5, pp. 94-115 (2011)
- [4] Y. Ke, X. Tang, and F. Jing: "The Design of High-Level Features for Photo Quality Assessment", Pproc. of IEEE Conference on Computer Vision and Pattern Recognition, pp. 419-426 (2006)
- [5] L. Kang, P. Ye, Y. Li, and D. Doermann: "Convolutional Neural Networks for No-Reference Image Quality Assessment", Proc. of IEEE Conference on Computer Vision and Pattern Recognition, pp. 1733-1740 (2014)
- [6] S. Bosse, D. Maniry, T. Wiegand, and W. samek: "A Deep Neural Network for Image Quality Assessment", Proc. of International Conference on Image Processing, pp. 3773-3777 (2016)
- [7] X. Jin, J. Chi, S. Peng, Y. Tian, C. Ye, and X. Li: "Deep Image Aesthetics Classification using Inception Modules and Fine-tuning Connected Layer", Proc. IEEE Wireless Communications & Signal Processing, pp. 1-6 (2016)
- [8] T. Hossein, and M. Peyman: "NIMA: Neural Image Assessment", arXiv:1709.05424 (2017)
- [9] F. Omori, H. Takimoto, H. Yamauchi, A. Kanagawa, T. Iwasaki, and M. Ombe: "Aesthetic Quality Evaluation Using Convolutional Neural Network", Asia-Pacific Journal of Industrial Management, Vol. VIII, Issue 1, pp. 71-77 (2019)
- [10] Microsoft Azure, "Face API - 顔認識ソフトウェア—Microsoft Azure", <https://azure.microsoft.com/ja-jp/services/cognitive-services/face/>
- [11] Google Cloud, "Vision API - 画像コンテンツ分析—Cloud Vision API—Google Cloud", <https://cloud.google.com/vision/?hl=ja>
- [12] F. Chollet: "Xception: Deep Learning with Depthwise Separable Convolutions", Proc. of IEEE Conference on Computer Vision and Pattern Recognition, pp. 1800-1807 (2017)
- [13] S. Christian, L. Wei, J. Yangqing, S. Pierre, R. Scott, A. Dragomir, E. Dumitru, V. Vincent, and R. Andrew: "Going Deeper with Convolutions", arXiv:1409.4842 (2014)
- [14] S. Christian, V. Vincent, I. Sergey, S. Jonathon, and W. Zbigniew: "Rethinking the Inception Architecture for Computer Vision", arXiv:1512.00567 (2015)
- [15] S. Christian, I. Sergey, V. Vincent, and A. Alex: "Inception-v4, Inception-ResNet and the Impact of Residual Connections on Learning", arXiv:1602.07261 (2016)
- [16] J. Deng, W. Dong, R. Socher, L-J. Li, and L. Fei-Fei: "ImageNet: A Large-Scale Hierarchical Image Database", Proc. of Computer Vision and Pattern Recognition (2009)
- [17] N. Murray, L. Marchesotti, and F. Perronnin: "AVA: A Large-Scale Database for Aesthetic Visual Analysis", Proc. of IEEE International Conference on Computer Vision and Pattern Recognition, pp. 2408-2415 (2012)



(a) 予測値:5.8018 (真値:5.8)



(b) 予測値:5.4689 (真値:5.56)



(c) 予測値:6.3164 (真値:6.7839)



(d) 予測値:4.6667 (真値4.2944)

図 6: 提案モデル (特徴次元数 293) による審美スコアの予測値 (真値) の例