生成顔の「お人形さん」っぽさの定量化

—— 左右対称性と平均性における実写との幾何学的比較 ——

Muse† ・ Hisashi Itoh†

† 個人　https://github.com/hisashi-ito　it.hisas@gmail.com（Hisashi Itoh）

Quantifying the “Doll-Like” Quality of AI-Generated Faces: A Geometric Comparison of Bilateral Symmetry and Averageness with Real Photographs

Muse† and Hisashi Itoh†

†Independent

概　要

生成モデルで作った実写系の顔写真は、破綻がないにもかかわらず「AI っぽい」「誰でもない顔」に見えることが多い。本稿はこの感覚を、不気味の谷ではなく、左右対称に整えられ個体差を削がれた「お人形さん」っぽさと捉え、顔ランドマークの幾何量で定量化する。生成系 2 群（API 系生成モデルとオープンウェイトモデル P-H3-Image による日本人女性の正面ポートレート、各 100 枚）と実写正面群（ライセンス明記の実在写真 100 枚）に同一の検出器（MediaPipe、478 点）を適用し、(1) 左右対応点の非対称量、(2) 平均形状への近さと群内の幾何分散、(3) 頬領域の肌テクスチャ（高周波エネルギー）を比較した（本比較）。非対称量は生成 2 群とも実写群より小さく（全体で Cohen’s $d = -0.92$、95%CI $[-1.20, -0.70]$ および $d = -1.11$、$[-1.37, -0.91]$）、全 7 部位でも同方向であった。自群平均形状までの距離でも生成群が小さく（$d = -1.78$ および -1.38）、姿勢を厳しく制限した感度分析でも結論は変わらなかった。一方、肌テクスチャは生成系に依存した。API 系の生成群は実写群より明確に滑らかであった（中央値 164 対 651、$d = -1.59$）のに対し、H3 群では差がなく（$d = +0.25$）、生成 2 群どうしは完全分離した。生成 2 群の間の非対称量には差が検出されなかった（$d = +0.18$、$p = 0.76$）。各群の平均形状そのものの非対称量は個体の平均を下回り（実写群では平均化により個体平均の 28% まで低下）、実写群と H3 群では平均形状への近さと非対称量の群内相関（$r \approx 0.57$）が姿勢を統制しても残ったことから、対称性は平均性の帰結として結びついていることを確認した。本稿に先行する小標本の第 1 回計測（生成 30 枚・実写 25 枚）でも対称性と平均性の方向は同じであった。先行研究が知覚評定と埋め込み空間で示してきた「過剰な平均性」のうち、対称性と平均性は生成系を問わない幾何量として実測でき、テクスチャは系と条件に依存する、というのが本稿の結論である。

キーワード：生成顔，顔ランドマーク，左右対称性，平均顔，効果量，MediaPipe

1. はじめに

画像生成モデルで作った実写系の人物ポートレートは、一目で「AI だ」と分かることがある。耳の形が崩れている、瞳がおかしい、といった破綻の時代は過ぎつつあり、現在の生成顔はむしろ破綻がない。それでも残る違和感は、不気味の谷（人間に近いが届かない造形への拒否反応）とは性質が違う。整っている。きれいである。しかし実在感がなく、「誰でもない誰か」に見える。本稿ではこの質を、左右対称に彫られ、個体差としての歪みを持たない人形の顔に例えて「お人形さん」っぽさと呼ぶ。

この感覚には先行研究の裏付けがある。Dawel ら [1] は、AI 生成顔が実在顔より「より対称的・比例的・魅力的だが、個性・記憶性・表情に乏しい」と評定されることを示し、この 6 特性への注意を訓練すると検出精度が約 41% から 81% に上がることを報告した。著者らの総称は hyper-average（過剰に平均的）である。Miller ら [2] は AI hyperrealism と呼ばれる現象、すなわち AI 顔が実在顔より「人間らしい」と誤認されやすいことを示し、対称性・平均性・肌の滑らかさが AI 顔側の特徴であることを評定で同定した。Nightingale と Farid[3] は、StyleGAN2 の顔について判別がチャンス水準（48.2%）にとどまり、合成顔の方が信頼性が高く評定されることを示した。幾何に近い測定としては、Dunn ら [4] が深層ニューラルネットの顔埋め込み空間で AI 顔が同性別の重心に有意に近いことを確認している。検出可能性は生成系譜に強く依存し、拡散系の顔は GAN 系より当てやすい（McGuire ら [5] は拡散 62.1% 対 GAN 48.0%）。

ただし、これらの知見は知覚評定か、解釈の難しい埋め込み空間での測定が中心である。ランドマークの幾何量そのもの —— 左右対応点の非対称量や、形状の群内分散 —— を、現行世代（拡散／ API 系）の生成顔と実在写真とで同一パイプラインで実測し分布比較した研究は、筆者らの調査では見当たらない。古典文献の側でも、平均性 [6] と対称性は魅力を「高める」要因とされており、「整いすぎ」がなぜ実在感の欠如につながるかは、実測なしには論じにくい。本稿の問いは次の 3 つである。

RQ1（対称性）生成顔の左右非対称量は、実写の顔より小さいか。
RQ2（平均性）生成顔の形状は群の平均形状の近くに集まり、幾何分散は実写より小さいか。
RQ3（テクスチャ）生成顔の肌は、高周波成分の少なさという意味で実写より滑らかか。

ここで、RQ1 と RQ2 の関係について、本稿の見立てを述べておく。実在の顔の左右差は、発生過程のノイズとして生じる微小な左右差、すなわち揺らぎ非対称であり、その方向は個体ごとにランダムである。方向がランダムな左右差は集団の平均では打ち消し合うため、平均顔そのものはほぼ左右対称になるはずである。生成モデルが学習分布の中心付近から顔を構成するなら、生成顔は平均形状の近くに集まり（RQ2）、その帰結として左右対称にも近づく（RQ1）はずである。つまり対称性と平均性は独立した 2 つの性質ではなく、平均性を主、対称性をその帰結とする結びつきが予想される。この見立ては、言われてみれば当たり前に聞こえるかもしれない。しかし「当たり前である」ことと、それが解釈可能な幾何量として実測でき、その大きさがどれほどであるかは別の問題であり、先行研究はこの実測を欠いていた。本稿では、この結びつき自体も検証の対象とし、各群の平均形状そのものの非対称量と、平均形状への近さと非対称量の群内相関を併せて報告する（3.7 節）。

本稿は 2 段階の計測を報告する。第 1 段階は、生成群 30 枚と実写群 25 枚の小標本比較（第 1 回計測）と、生成群を 100 枚に増やした予備的計測である。第 2 段階が本稿の主結果であり、生成系 2 群と、監査基準を最初から適用して再構築した実写正面群との、$n = 100$・3 群の比較（本比較）である。第 1 回計測の対照群は統制が弱く、その効果量には撮影条件の差が混ざり得た。本比較は、その確定条件として用意したものである。なお本稿は生成顔の優劣や真贋判定の実用を論じるものではなく、違和感の正体と目される幾何学的性質が実在するか、その大きさはどれほどかを測ることが目的である。

2. 方法

表 1　サンプル群の構成

群	n	由来と条件
生成群	100	API 系生成モデル。正面・頭部〜肩・無地背景。年齢層 4 層を回す（第 1 回は 30 枚）
H3 群（v2）	100	オープンウェイトモデル P-H3-Image。拡張プロンプト固定、41 枠は姿勢基準で再生成し差し替え
実写正面群	100	Wikimedia Commons の CC 写真。映画祭レッドカーペットが中心。監査基準で選抜
旧実写群	25	CC ／ PD の実在写真 27 枚のうち検出成功分。第 1 回計測のみに使用

2.1 サンプル群

サンプル群の構成を表1に示す。本比較は生成系 2 群と実写正面群の 3 群（各 100 枚）からなる。

生成群（本稿では Muse 群とも呼ぶ）は、API 系の画像生成モデル（Muse の画像生成パイプライン）で作った日本人女性の実写系ポートレートである。プロンプトは骨格を固定し（正面顔、頭部から肩まで、カメラ目線、柔らかい自然な表情、無地の室内背景、柔らかい自然光）、年齢層の語句のみを 20 代前半・後半、30 代前半・後半の 4 層で回した。解像度は 1600×1600、形式は WebP である。全枚を目視し、正面でないもの、複数人が写るもの、顔の破綻があるものを除外する基準で選抜した。本比較では 100 枚（年齢層は各 25 枚）を用い、第 1 回計測ではそのうち 30 枚を用いた。なお年齢層のラベルの一部は、生成記録からの再構成を含む。

第 2 の生成系は、オープンウェイトの画像モデル P-H3-Image（Pruna AI、MiniMax H3 系譜）による生成群（H3 群）である。生成は第 2 著者の自宅計算機上で行った。生成条件の実績は次のとおりである。量子化は int8_convrot（モデル本体と文章エンコーダ）、VAE は bf16、解像度 2048×2048（PNG）、8 ステップ（euler）、蒸留モデルのため CFG とネガティブプロンプトはなし。プロンプトは公式ワークフローの拡張器による拡張を有効とし、生成群と同じ骨格文を入力した。拡張は貪欲法のため、拡張後の文は年齢層ごとに 1 つに固定され、同じ層の 25 枚は拡張文が同一でシードのみが異なる。各年齢層で 30 枚を生成し、目視で 25 枚に選抜した。ただし到着後のこちらの計測で、100 枚中 41 枚が姿勢の基準（後述のパイロットで上限とした実測値、ヨー代理量の絶対値 0.098）を超えていたため、該当する 41 枠については、同一条件・同一の拡張文のままシードだけを変えて再生成し（41 枠で候補 364 枚）、生成側で MediaPipe により最も正面の候補を選んで差し替えた。差し替え後の群を H3 群（v2）と呼ぶ。選抜に使ったヨーの定義は生成側（顔の変換行列から求めたラジアン、1024×1024 に縮小して計測）と計測側（ランドマーク形状からの代理量、原寸で計測）で異なり、絶対値は直接は対応しない。差し替えの効果は計測側の代理量でも確認できた（群のヨー代理量の中央値が 0.070 から 0.033 へ、非対称量の平均が 0.0513 から 0.0233 へ低下）。

実写群は 2 種類ある。第 1 回計測の実写群は、ライセンスが明確で研究目的の内部利用に適合する実在写真で、出典は Flickr のクリエイティブ・コモンズ写真（Openverse 経由）と Wikimedia Commons のパブリックドメイン画像である。1 枚ごとに出典 URL とライセンスを取得記録（sources.csv）に残した。日本人女性、正面寄り、顔が十分に大きいものを目視で選んだ。27 枚を集め、後述の検出に成功した 25 枚を分析対象とした。収集後の属性監査で、この群は統制が弱いことが分かった（2.5 節と限界）。

本比較の対照群である実写正面群は、この監査で得た除外基準を最初から適用して再構築した群である。出典は Wikimedia Commons の CC 写真で、東京国際映画祭・横浜国際映画祭のレッドカーペット写真が中心である（CC BY 2.0 が 99 枚、CC BY-NC-ND 2.0 が 1 枚）。女性のみとし、10 代に見えるもの、モノクロ・ヴィンテージ、閉眼、明らかに非正面のもの、顔が小さい・切れているものを除外し、同一人物については最良の 1 枚のみを残した。計 100 枚を収集し、検出は全数成功した。被写体は女優等の公人で、プロの撮影による、メイク済みの写真である。年齢層の統制はない。いずれの実写群についても、写真そのものは公開せず、計測は手元の環境内でのみ行った。なお本稿の図版方針として、掲載する顔写真は合成画像（生成群・H3 群）のみとし、実写サンプルの写真は、被写体の肖像権および CC ライセンスの条件に配慮して掲載しない。実写群については、数値とランドマーク点群のみを示す。

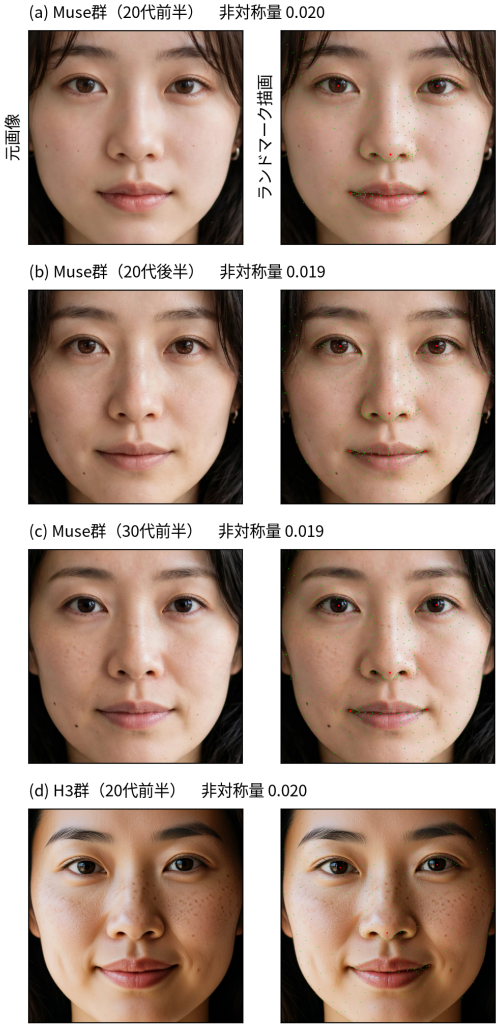


図 1　生成サンプルの例。各行は左が元画像、右がランドマーク描画（478 点、アンカー点を赤で強調）。(a)–(c) は Muse 群から、非対称量（全体）が群の中央値（0.019）付近のものを年齢層を変えて選んだ（順に 20 代前半・20 代後半・30 代前半）。(d) は H3 群（v2、20 代前半）から中央値（0.020）付近のものを選んだ（非対称量 0.020）。実写サンプルは本稿に掲載しない（2.1 節）

2.2 ランドマーク取得と検証

顔ランドマークは MediaPipe Face Landmarker（478 点＝ 468 点メッシュ＋虹彩 10 点、float16 モデル）で取得した。全群に同一の検出器・同一条件を適用した。生成群のサンプルとランドマーク描画の例を図1に、各群の平均形状（ランドマーク点群のみ）を図2に示す。

測定の妥当性は描画で検証した。各画像にランドマークを重ねた描画画像を作り、虹彩中心・鼻尖・口角・顎・額などのアンカー点が解剖学的位置に乗っているかを全数（本比較の 3 群は各 100 枚、第 1 回の実写群は検出成功 25 枚、H3 パイロット 4 枚、H3 差し替え分 41 枚）目視した。いずれの群でも、系統的なずれ、メッシュの崩壊、別人への誤検出は認められず、描画検証を理由に除外した画像はない。実写正面群では、画面の端や背景に他人が写る枚でも検出は中央の主被写体を選んでおり、口元にマイクを持つ枚や、強い笑顔でほぼ閉眼の枚でもメッシュは顔に乗っていた。過去の同様の測定で、格子当てはめの誤りから結論を撤回した経験があるため、「検出できたこと」と「検出が正しいこと」は分けて確認する手順を取った。

2.3 検出器の適合性

検出器（MediaPipe）そのものの適合度を、4 つの観点で定量化した（表2）。

第 1 に検出成功率である。生成群は 100 枚中 100 枚、H3 パイロットは 4 枚中 4 枚で検出に成功した。実写群は 27 枚中 25 枚で、失敗した 2 枚は、明治期の写真（台紙に貼られた小さなプリントでセピア・低コントラスト、顔が非常に小さい）と、大きく横を向いた半横顔（片目がほぼ隠れる）であり、いずれも本計測の対象条件（現代のカラー写真でほぼ正面）から外れた画像である。なお本比較の群でも検出は全数成功した（H3 群 100 枚、実写正面群 100 枚。い

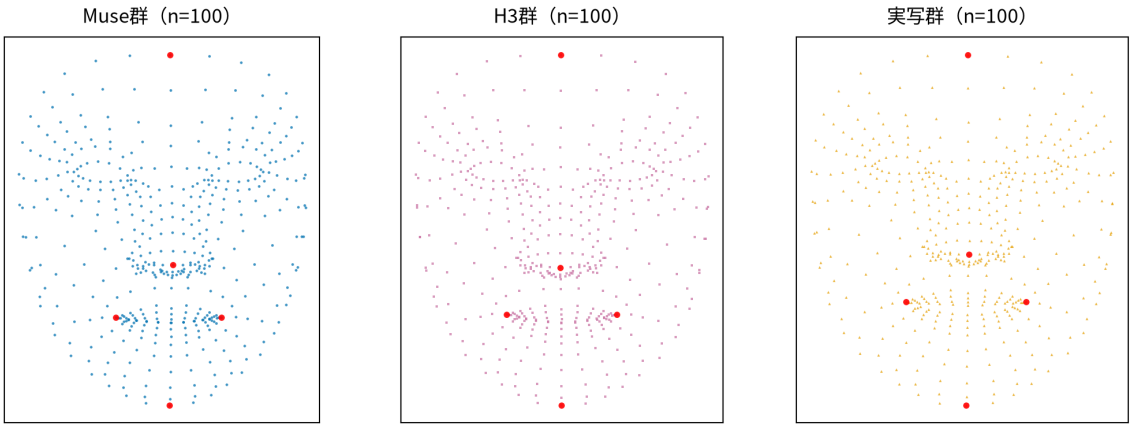


図 2　本比較の各群の平均形状（各 $n = 100$）。相似正規化（両目中心の中点を原点、目間距離を 1）した 468 点形状を Procrustes 整列して平均したランドマーク点群であり、写真は含まない。赤はアンカー点（額・鼻尖・口角・顎）。実写群は点群のみであれば被写体を特定する情報にならないため、この形でのみ図示する

ずれも失敗なし）。以下の第 2〜第 4 の評価は、生成群・第 1 回の実写群・H3 パイロットを材料に行ったものである。

第 2 に再実行の決定性である。各群から 39 枚を再検出し、保存済み座標との最大絶対差は 5.0×10^{-7} であった。これは座標を小数 6 桁で記録した際の丸め誤差の範囲であり、実質的に完全一致とみなせる。なお、パイプラインの文書には「最大絶対差 0」と記したが、厳密にはこの丸めの範囲の差があるため、本稿では実測値をもって訂正する。

第 3 に描画の全数目視である（前節）。生成群 100 枚については、頭部の軽い傾きが見られた数枚でもメッシュは顔に追随しており、検出誤りと判定した画像はなかった。

第 4 に、左右反転テストを行った。各画像を水平反転して再検出し、反転検出の座標を鏡映し、解剖学的な左右ラベルを対応ペア表で付け替えて元の座標と比べた（MediaPipe の点番号は解剖学的な左右を表すため、この付け替えが要る）。全 478 点のずれの RMS を目間距離で割った値の中央値は、生成群 0.022、実写群 0.028、H3 パイロット 0.023 で、群間に大きな差はない。ただし実写群はばらつきが大きく（SD 0.020）、ずれが最大の画像（0.117）は顔が小さく画面端で切れた夜景の写真であった。生成・H3 群のずれ（0.023 前後）は、検出器が顔そのものに対して報告する鏡映距離の RMS（生成群では非対称量 0.026 に対応する約 0.040）の約 6 割に達する。つまり、反転で分かる範囲の検出揺らぎは、生成群の非対称量と同じ桁にある。このことは、今回の対実写比較（非対称量の差は約 0.08）の結論には影響しないが、非対称量が小さい群どうしの比較（第 2 の生成系との本計測など）では、反転ずれが無視できない誤差要因になる。実写群では、反転した 1 枚（全身を俯瞰した非正面の写真）で検出自体が失敗した。また反転画像から算出した非対称量（全体）は、元の値と強く相関した（生成群 $r = 0.96$、実写群 $r = 0.99$、H3 $r = 1.00$）。絶対差の平均は生成群 0.0047、実写群 0.0096 であった。画像を左右反転しても非対称量の順位も大きさもほぼ保たれることから、本稿の非対称量は検出器の片寄り（左右非対称な検出癖）で作られた人工物ではない、とまでは言える。ただし、反転前後の非対称量の平均には小さな系統差が見られた（生成群では反転後が平均 $+0.0018$、実写群では -0.0033）。向きに依存する検出の揺らぎが残るため、非対称量の絶対値の解釈には幅を持たせるべきであり、群間差がこの揺らぎの範囲内にある生成群どうしの比較では、反転した画像との平均などの対策を検討する。なお、両眼が同一人物のものであることを利用する検証（眼球反射の左右一致など）は本稿では実施していない。

2.4 指標

非対称量は次のように定義した。両目の中心が水平になるよう回転し、目間距離が 1 になるよう相似正規化する。正中線で顔を鏡映し、左右対応点ペアの間の距離の平均を非対称量とする。対応ペア（220 ペアと正中線上の 28 点）は、MediaPipe の公式な標準顔モデル（頂点配置が厳密に左右対称）から、鏡映と相互最近傍で機械的に導出した。手作業で書いたペア表は誤りを含むことが検証で分かったため破棄し、推測に依存しないこの導出に置き換えた。全体のほか、部位別（目・眉・鼻・頬・口・輪郭・下顔面）の 7 種も算出した。

幾何分散は、相似正規化した 468 点形状を Procrustes 平均形状へ整列させたうえで、(i) 各画像の自群平均形状までの距離（RMS）、(ii) 両群統合の平均形状までの距離、(iii) 群レベルの点ごと群内標準偏差の平均と群分散（二乗距離の平均）を算出した。(ii) は「平均顔への近さ／個性のなさ」の代理指標である。

肌テクスチャは、頬領域の高周波エネルギーで定義した。元画像を目間距離 300 ピクセルにリサイズし、統合平均形状上で頬の外側 1/3 に位置する点から 3×3 グリッドで選んだ固定 18 アンカーに、中心を顔中心側へ 30% 寄せた 90×90 ピクセルのパッチを置き、各パッチの Laplacian 分散の平均を取った。アンカーを固定したのは、ランドマークの密度が鼻周辺に偏るため、パッチ位置が画像ごとにずれると測定部位が変わってしまうからである。

姿勢の診断量として、ヨーの代理量（鼻尖の両目中点からの水平ずれを目間距離の半分で割った値の絶対値）と、ロール（目線の傾き、度）を併せて記録した。

2.5 統計と監査

群間比較では、群ごとの平均・中央値・標準偏差に加え、Cohen's d（生成群-実写群、プールした標準偏差）、Mann–Whitney の U 検定（両側）の p 値、AUC（生成群の値が実写群より大きい確率）を併記した。予備的計測と本比較では d にパーセンタイル・ブートストラップによる 95% 信頼区間（再標本化 2,000 回）を付けた。p 値より効果量を主役とする。

第 1 回計測の実写群については、収集後に属性監査を行った。全 27 枚を目視し、性別・年代の見た目・正面度・生成群との条件適合（適合／要注意／不適合）を記録した。結果は限界の節で述べる。この監査で得た除外基準（10 代に見えるもの、同一人物の重複、モノクロ・ヴィンテージ、閉眼、明らかな非正面、顔が小さいもの）は、本比較の実写正面群の収集に最初から適用した。

3. 結果

3.1 本比較（3 群・各 $n = 100$）

本比較の主結果を表3に、分布を図3に、対実写の効果量を図4に示す。検出は 3 群とも全数成功した（各 100 枚、失敗なし）。姿勢の診断量（ヨー代理量の絶対値）の中央値は、Muse 群 0.024、H3 群 0.033、実写正面群 0.058 であった。ただし実写正面群は前後半で質が違う。前半 50 枚は中央値 0.025・最大 0.071 と生成群並みに正面であるのに対し、後半 50 枚は中央値 0.110 とやや緩く、0.2 を超えるものが 22% あった（最大 0.267）。3 群とも 0.3 を超える画像はなかった。この偏りが結果を作っていないかは、感度分析（3.6 節）で確認する。

3.2 RQ1：対称性

非対称量（全体）は、Muse 群の平均 0.0264（中央値 0.0193）、H3 群 0.0233（同 0.0202）に対し、実写正面群 0.0542（同 0.0421）で、実写は生成の約 2 倍である。効果量は Muse 群-実写群で $d = -0.92$（95%CI $[-1.20, -0.70]$、$p = 1.0 \times 10^{-12}$、AUC 0.208）、H3 群-実写群で $d = -1.11$（$[-1.37, -0.91]$、$p = 6.2 \times 10^{-15}$、AUC 0.181）。部位別でも 7 部位すべてで生成側が小さく、最大は口（Muse 群 -1.16、H3 群 -1.07）と輪郭（同 -1.04、-0.99）であった。第 1 回計測（全体 $d = -1.43$）と比べ効果量は縮小したが、これは対照群がスナップ中心の群から正面・プロ撮影の統制の取れた群に置き換わった結果と読める。統制を強めても大きな効果が残り、部位の順序（口が最大）も再現した。RQ1 は本比較でも支持された。

表 2　検出器（MediaPipe Face Landmarker）の適合性評価。反転ずれは「反転検出を鏡映・ラベル付け替えして元座標と比べた全 478 点の RMS（目間距離比）」

項目	生成群	実写群	H3 パイロット
検出成功／総数	100/100	25/27	4/4
描画目視（系統的ずれ）	100 枚全数、なし	25 枚全数、なし	4 枚全数、なし
再実行の最大絶対差	39 枚で 5.0×10^{-7}（記録の丸め誤差の範囲）		
反転検出の成功	100/100	24/25	4/4
反転ずれ RMS（中央値）	0.022	0.028	0.023
反転ずれ RMS（平均 ±SD）	0.023 ± 0.002	0.033 ± 0.020	0.023 ± 0.002
非対称量の反転前後相関 r	0.96	0.99	1.00
非対称量の絶対差の平均	0.0047	0.0096	0.0057

表 3　本比較の主結果（Muse 群・H3 群・実写正面群、各 $n = 100$）。平均 ±SD と、対実写の Cohen's d（生成群-実写群、括弧内はブートストラップ 95% 信頼区間）

指標	Muse 群	H3 群	実写群	d（Muse-実写）	d（H3-実写）
非対称・全体	0.0264 ± 0.0205	0.0233 ± 0.0127	0.0542 ± 0.0373	-0.92 $[-1.20, -0.70]$	-1.11 $[-1.37, -0.91]$
非対称・目	0.0059 ± 0.0032	0.0051 ± 0.0024	0.0086 ± 0.0046	-0.67 $[-0.95, -0.41]$	-0.96 $[-1.24, -0.72]$
非対称・眉	0.0225 ± 0.0155	0.0173 ± 0.0094	0.0373 ± 0.0260	-0.69 $[-0.96, -0.45]$	-1.02 $[-1.27, -0.82]$
非対称・鼻	0.0435 ± 0.0421	0.0335 ± 0.0282	0.0780 ± 0.0620	-0.65 $[-0.93, -0.40]$	-0.92 $[-1.18, -0.70]$
非対称・頬	0.0292 ± 0.0234	0.0238 ± 0.0143	0.0541 ± 0.0385	-0.78 $[-1.06, -0.54]$	-1.04 $[-1.30, -0.84]$
非対称・口	0.0184 ± 0.0119	0.0211 ± 0.0108	0.0514 ± 0.0384	-1.16 $[-1.46, -0.98]$	-1.07 $[-1.34, -0.90]$
非対称・輪郭	0.0283 ± 0.0212	0.0313 ± 0.0165	0.0779 ± 0.0643	-1.04 $[-1.29, -0.86]$	-0.99 $[-1.24, -0.82]$
非対称・下顔面	0.0241 ± 0.0197	0.0213 ± 0.0152	0.0495 ± 0.0349	-0.89 $[-1.17, -0.65]$	-1.05 $[-1.28, -0.82]$
自群平均形状までの距離	0.0192 ± 0.0098	0.0249 ± 0.0084	0.0417 ± 0.0150	-1.78 $[-2.15, -1.49]$	-1.38 $[-1.69, -1.12]$
固定参照形状までの距離	0.0165 ± 0.0086	0.0242 ± 0.0077	0.0502 ± 0.0159	-2.63 $[-3.09, -2.29]$	-2.08 $[-2.47, -1.77]$
肌テクスチャ	170 ± 47	890 ± 269	784 ± 545	-1.59 $[-1.89, -1.38]$	$+0.25$ $[-0.03, +0.59]$

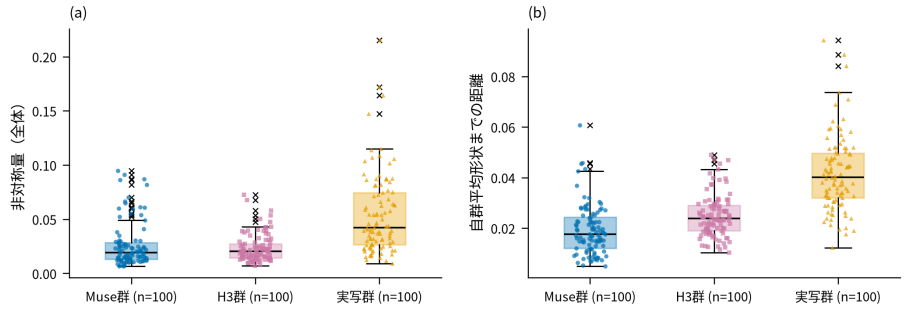


図 3　本比較における分布（各群 $n = 100$）。(a) 非対称量（全体）。(b) 自群平均形状までの距離。箱は四分位範囲、点は個々の画像。生成 2 群はどちらの指標でも低い側に固まっている

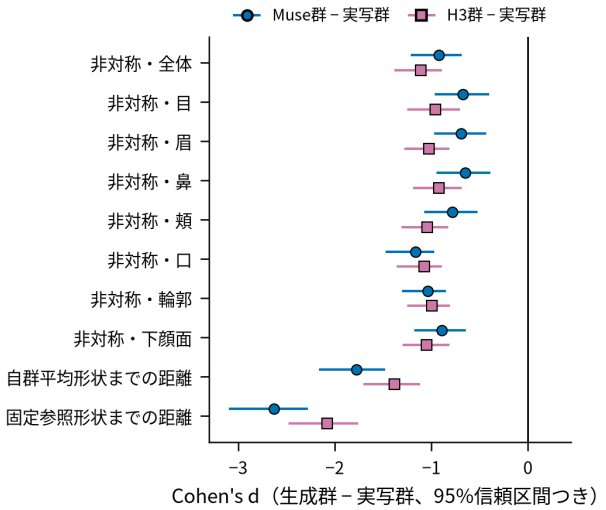


図 4　本比較の効果量（Cohen's d、生成群-実写群）。点が d、線がブートストラップ 95% 信頼区間。幾何系の全指標が負（生成群が小さい）方向に大きく、2 つの生成系でよく揃っている

3.3 RQ2：平均性

自群平均形状までの距離は、Muse 群 0.0192、H3 群 0.0249、実写正面群 0.0417 で、効果量は $d = -1.78$（$[-2.15, -1.49]$）および -1.38（$[-1.69, -1.12]$）と、本比較で最大級である。点ごと標準偏差の平均も実写群 0.0291 に対し Muse 群 0.0140、H3 群 0.0172 と、実写は生成の約 1.7〜2.1 倍である。群分散（二乗距離の平均）でも、実写群 1.97×10^{-3} に対し Muse 群 4.65×10^{-4}（実写の約 1/4）、H3 群 6.90×10^{-4}（約 1/3）であった。固定参照形状までの距離でも生成群が小さい（$d = -2.63$ および -2.08）。ただし参照形状は生成群（Muse 群）100 枚と旧実写群 25 枚の統合平均であり、Muse 群自身がプールに含まれる分、Muse 群の値は機械的に小さくなる。プールの外にある H3 群と実写群の比較（-2.08）は、その割り引きなしに読める。RQ2 は本比較でも支持された。

3.4 RQ3：テクスチャ

肌テクスチャは、生成系によって結論が分かれた。平均（中央値）は Muse 群 170（164）、H3 群 890（857）、実写正面群 784（651）である。Muse 群-実写群では $d = -1.59$（$[-1.89, -1.38]$、AUC 0.043）と、生成側が明確に小さい。この対照群では「生成顔の肌は滑らか」が成り立つ。対照的に H3 群-実写群では $d = +0.25$（$[-0.03, +0.59]$、$p = 4.6 \times 10^{-4}$）と差はなく、方向も逆である。生成 2 群どうしは $d = -3.72$（$[-4.31, -3.30]$）、AUC 0.000 と完全分離した。第 1 回計測でテクスチャが有意でなかったのは、旧実写群の中央値が 179.5 と低く、対照側の写真条件が生成側に近かったためと判断できる。H3 群の拡張プロンプトは毛穴・そばかす・フィルム調の粒を明示的に要求しており（2.1 節）、テクスチャの水準はモデルそのものよりプロンプトと条件に強く依存する。RQ3 は、生成系一般の性質としては支持されず、系と条件に依存する。

3.5 生成群どうしの比較と年齢層別

生成 2 群どうしでは、非対称量（全体）に差が検出されなかった（$d = +0.18$、95%CI $[-0.09, +0.43]$、$p = 0.76$、AUC 0.487）。部位

表 4　感度分析（姿勢制限下の効果量 d、生成群-実写群）。残数：制限なし各 100、≤ 0.2 で Muse 98・H3 99・実写 89、≤ 0.1 で Muse 87・H3 97・実写 74

指標（比較）	制限なし	≤ 0.2	≤ 0.1
非対称・全体（Muse-実写）	-0.92	-0.92	-1.15
非対称・全体（H3-実写）	-1.11	-1.07	-1.02
自群平均距離（Muse-実写）	-1.78	-1.86	-2.19
自群平均距離（H3-実写）	-1.38	-1.35	-1.46
肌テクスチャ（Muse-実写）	-1.59	-1.61	-1.66
肌テクスチャ（H3-実写）	$+0.25$	$+0.22$	$+0.15$

別も効果量の絶対値は最大で 0.41（眉）にとどまり、方向も混在していて、系統的な差とは読めない。H3 群の当初到着分（差し替え前）ではこの比較が $d = -0.74$ であったが、その差は姿勢の差であり、姿勢を統制した差し替えの後に消えた。形状の分散は H3 群がやや大きい（自群平均形状までの距離 $d = -0.62$、固定参照形状まで $d = -0.95$）。テクスチャは前節のとおり完全分離した。

年齢層別の非対称量（全体）の平均は、Muse 群が 20 代前半 0.0263、後半 0.0294、30 代前半 0.0293、後半 0.0205、H3 群が同 0.0194、0.0301、0.0197、0.0240 であった。同層どうしの効果量は $+0.53$ から -0.22 の範囲で信頼区間は広く、層のパターンと呼べるものはない。なお H3 群では拡張文が層ごとに固定され、髪型・服・背景も層ごとに揃っているため、H3 群の層間比較には年齢以外の差が交絡する（限界の節）。

3.6 感度分析

姿勢の交絡への感度分析として、全群にヨー代理量の制限を課して再計算した（表4）。制限 ≤ 0.3 では除外がゼロのため、主分析と同一である。≤ 0.2（実写 89 枚が残る）では非対称・全体の d は -0.92 ／ -1.07、≤ 0.1（実写 74 枚）では -1.15 ／ -1.02 と、姿勢を厳しくするほど Muse 群側はむしろ強まった。自群平均形状までの距離とテクスチャでも、制限による結論の変化はなかった。実写正面群の後半採用分の振れが、本比較の結果を作っているわけではない。生成群どうしの比較は ≤ 0.1 で $d = -0.29$（$[-0.56, 0.00]$、$p = 0.076$）とやや動いたが、有意水準には届かず、差なしという結論は変わらない。

3.7 平均性と対称性の結びつき

はじめにで述べた見立て（平均性が主で、対称性はその帰結である）を、本比較のデータで直接確かめる分析を 2 つ行った（表5）。

第 1 に、各群の平均形状そのものに非対称量の定義を適用した。平均形状の非対称量（全体）は、Muse 群 0.0229、H3 群 0.0097、実写正面群 0.0150 で、いずれも個体の非対称量の平均（順に 0.0264、0.0233、0.0542）を下回った。とくに実写正面群では、平均化によって非対称量が個体平均の 28% まで低下した。個体の左右差の大部分は方向がランダムな揺らぎ非対称であり、平均で打ち消し合うことの直接の実測である。実写群の平均形状（0.0150）は生成 2 群の個体の平均（0.0264、0.0233）よりも対称であり、「平均する」という操作だけで、生成個体より対称な顔が得られる。一方、Muse 群では平均形状の非対称量が個体平均の 87% にとどまった。Muse 群には符号付きヨー代理量に一方向の偏り（平均 $+0.043$、93% が同符号）があり、群で共有された弱い顔の振れは平均しても残る。Muse 群の小さな非対称量は、個体ごとの揺らぎよりも、こうした群共通成分で主に構成されていると読める。

第 2 に、群内での平均形状への近さと非対称量の相関を見た。Pearson 相関は Muse 群 $r = 0.585$、H3 群 $r = 0.578$、実写正面群 $r = 0.570$（いずれも $p < 10^{-9}$）と、3 群とも同程度であった。ただし内訳は群で異なる。実写群と H3 群では、この相関はヨー代理量の絶対値を統制しても残った（偏相関 0.567 および 0.548。ヨー代理量の絶対値が 0.1 以下の画像に限っても $r = 0.547$ および 0.555）。これらの群では、平均形状に近い個体ほど対称であるという結びつきが、個体レベルでも観測される。対照的に Muse 群の相関は姿勢が駆動していた。Muse 群ではヨー代理量と非対称量の相関が 0.94 と高く、ヨー代理量の絶対値が 0.1 以下の画像に限ると相関は消え（$r = -0.10$）、偏相関も -0.07 であった。Muse 群では、非対称量の個体差の主因は平均形状からの距離ではなく姿勢であり、平均性との結びつきは、自群平均形状までの距離そのものが小さいという、群の均質さとして現れている。

以上を合わせると、群レベルの読みは見立てと整合する。生成顔が対称なのは、対称性が独立に高められているためではなく、生成顔が集まる先である平均形状そのものが対称であり、生成顔はその近くに分布しているためと読める。対称性は平均性の帰結である。

表 5　平均性と対称性の結びつき（本比較、3 群とも $n = 100$）。平均形状の非対称量は、各群の Procrustes 平均形状そのものに非対称量（全体）の定義を適用した値。相関は、自群平均形状までの距離と非対称量（全体）の群内相関

群	個体（平均）	平均形状	比	Pearson r	Spearman ρ
Muse 群	0.0264	0.0229	0.87	0.585	0.127
H3 群	0.0233	0.0097	0.42	0.578	0.417
実写正面群	0.0542	0.0150	0.28	0.570	0.496

3.8 先行する小標本の計測

本比較に先立つ 2 つの計測を、経緯として簡潔に残す。

第 1 回計測（生成群 30 枚、旧実写群 25 枚）では、非対称量は全 8 指標で生成群が小さく（全体 $d = -1.43$、最大は口の -1.77）、自群平均形状までの距離でも $d = -1.55$、群分散は実写の約 1/6 であった（表6）。テクスチャは有意でなかった（$p = 0.32$。平均は実写側の外れ値に引っ張られ、中央値は 179.5 対 159.5 とほぼ同じ）。ヴィンテージ枠の実写を除いた感度分析（実写 24 枚、非対称・全体 $d = -1.41$、口 -1.74、テクスチャ $p = 0.44$）や、描画目視で明らかに非正面と判定した 6 枚を除いた探索的分析（実写 19 枚、非対称・全体 $d = -1.26$、口 $d = -1.97$）でも、幾何系の結論は変わらなかった。ただしこの対照群は、後の属性監査で統制の弱さが判明した群である。監査（全 27 枚の目視）では生成群の条件への適合は 1 枚のみで、要注意 10 枚、不適合 16 枚であり、10 代に見えるもの 2 枚、同一人物の重複 2 組（実質 25 名分）、ヴィンテージ写真 4 枚、モノクロ 3 枚、閉眼 1 枚を含んでいた。効果量の絶対値は割り引いて読む必要があり、本比較で効果量が縮小したのは、この割り引きの範囲内である。

予備的計測（生成群 100 枚、旧実写群 25 枚）では、非対称量の方向は全指標で保たれ、効果量は増大して見えた（全体 $d = -2.07$、95%CI $[-2.81, -1.50]$、口 $d = -2.63$、表7）。生成群 100 枚の非対称量（全体）は平均 0.0264（SD 0.0205）であった。肌テクスチャは、平均の差から計算した d は -0.72 である一方、順位検定では $p = 0.29$ と有意でなかった。この予備的計測は対照側が統制の弱い旧実写群のままであり、効果の増大は対照群の質の問題を含む。本比較の値をもって置き換える。

4. 考察

本比較の結果は、「お人形さん」っぽさの少なくとも幾何学的な部分が、対照群の条件を変えても残る実測可能な性質であることを示した。生成顔は実在の顔より左右対称であり（RQ1）、互いの形状がよく似て平均顔の近くに集まっている（RQ2）。この 2 点は、生成系 2 群の双方で、正面・プロ撮影という統制の取れた対照群に対して大きな効果量で立った。Dawel ら [1] が知覚評定で得た「対称的・比例的だが個性に乏しい」という像、Dunn ら [4] が埋め込み空間で確認した「重心への近さ」は、本稿では解釈可能な幾何量として、大きな効果量を伴って再現された。生成モデルが学習分布の中心付近から顔を構成する以上、この方向の結果自体は機構的には自然である。自然であることと、測って数字で置くことは別の仕事であり、先行研究で空白だったのは後者である。

第 1 回計測からの変化も記録しておく。非対称・全体の効果量は、第 1 回の -1.43 から本比較の -0.92 ／ -1.11 へ縮小した。対照群の統制を強めれば効果が縮むのは、交絡が剥がれた分として自然であり、縮小した残りが、現時点でのこの現象の大きさの見積もりである。逆にテクスチャは、第 1 回では有意でなかったものが、本比較の Muse 群では明確な差になった。効果量の絶対値は対照群の作り方に依存する、という方法論的な教訓が、2 つのラウンドの対比から得られる。

古典文献との緊張関係に触れておく。平均性と対称性は、古典的には魅力を高める要因である [6, 8]。形状だけを操作してテクスチャを保てば完全対称の顔が好まれるという修正もなされている [7]。さらに、魅力的な平均顔を誇張した顔は単なる平均顔より好まれる [9]。つまり「整いすぎ」は「醜さ」ではない。本稿の結果が示すのは、魅力の欠如ではなく個体の不在である。実在の顔には、発生過程のノイズとして生じる微小な左右差（揺らぎ非対称）が必ず入り、それが「この個体である」ことの署名として働く。生成顔にはその署名が薄い。知覚が違和感として拾っているのは、欠陥ではなく、署名の欠如なのではないか。人形の顔が「誰の顔でもない」ことと同型である。

この「署名の欠如」は、3.7 節の分析でさらに具体化できる。実写群では平均化によって非対称量が個体平均の 28% まで消えた。すなわち実在の顔の左右差の大部分は、個体ごとに方向の違う揺らぎであり、平均すると残らない。それにもかかわらず実在の個体は、そ

表 6　第 1 回計測の主結果（生成群 $n = 30$、実写群 $n = 25$）。d は Cohen's d（生成群-実写群）、p は Mann–Whitney の両側 p 値、AUC は生成群の値が実写群より大きい確率

指標	生成平均	生成 SD	実写平均	実写 SD	d	p
非対称・全体	0.0282	0.0221	0.1117	0.0835	−1.43	3.5×10^{-6}
非対称・目	0.0063	0.0034	0.0166	0.0119	−1.24	5.5×10^{-5}
非対称・眉	0.0241	0.0172	0.0723	0.0568	−1.19	2.1×10^{-4}
非対称・鼻	0.0460	0.0449	0.1721	0.1623	−1.10	2.5×10^{-4}
非対称・頬	0.0311	0.0255	0.1105	0.0905	−1.25	2.9×10^{-5}
非対称・口	0.0199	0.0117	0.1042	0.0698	−1.77	7.1×10^{-9}
非対称・輪郭	0.0311	0.0225	0.1606	0.1242	−1.52	8.8×10^{-8}
非対称・下顔面	0.0255	0.0207	0.1079	0.0860	−1.38	2.7×10^{-6}
自群平均形状までの距離	0.0193	0.0111	0.0487	0.0253	−1.55	5.9×10^{-7}
統合平均形状までの距離	0.0159	0.0083	0.0605	0.0286	−2.21	1.7×10^{-9}
肌テクスチャ	165.4	39.1	372.5	628.1	−0.49	0.32

表 7　予備的計測（生成群 $n = 100$、実写群 $n = 25$）の効果量と 95% 信頼区間（ブートストラップ）

指標	d	95%CI	p
非対称・全体	−2.07	[−2.81, −1.50]	4.4×10^{-9}
非対称・目	−1.79	[−2.47, −1.25]	4.2×10^{-7}
非対称・眉	−1.73	[−2.42, −1.15]	2.1×10^{-6}
非対称・鼻	−1.59	[−2.21, −1.04]	5.8×10^{-6}
非対称・頬	−1.80	[−2.47, −1.23]	1.6×10^{-7}
非対称・口	−2.63	[−3.44, −2.11]	8.3×10^{-13}
非対称・輪郭	−2.28	[−3.06, −1.75]	4.3×10^{-11}
非対称・下顔面	−2.00	[−2.68, −1.50]	2.1×10^{-9}
自群平均形状までの距離	−2.07	[−2.77, −1.52]	1.3×10^{-9}
固定参照形状までの距離	−3.64	[−4.82, −3.00]	3.1×10^{-14}
肌テクスチャ	−0.72	[−1.28, −0.29]	0.29

の揺らぎを保持している。生成顔に薄いのは、まさにこの「平均すると消えてしまうはずの、個体固有の方向を持った左右差」である。生成モデルが分布の中心付近を構成する以上、この欠如は機構としては当たり前に思える。しかし、平均形状そのものの対称性を個体と同列に測り、平均への近さと非対称量の結びつきを群内で確かめる作業は、先行研究では行われていなかった。当たり前に思える帰結を、効果量と信頼区間を伴う数字として置くことが、本稿の測定としての寄与である。

部位別で最も効果が大きかったのは、本比較でも口であった。ただし実写正面群はレッドカーペットの、笑顔中心の写真であり、生成群は無表情に近い条件で統一されている。口の非対称には表情筋の使い方の左右差や撮影時の表情が強く乗るため、この部位の差の一部は表情条件の差である可能性が残る。部位別の解釈ではこの割り引きが要る。

RQ3（テクスチャ）については、本稿は当初の読みを改訂する。第 1 回で差が検出できなかった主因は、旧実写群の写真条件にあった（中央値 179.5 と、生成群並みに低かった）。本比較の対照群では Muse 群の滑らかさが明確に立ち（中央値 164 対 651）、他方で H3 群では立たず、生成 2 群どうしは完全に分かれた。つまり「生成顔の肌は滑らか」は生成系一般の性質ではなく、系とプロンプトに依存する性質である。H3 群の拡張プロンプトが毛穴・そばかす・フィルム調の粒を明示的に要求していたことは、この依存性の具体例である。なお「滑らかさ」はジャンルにも依存し得る。本稿の条件はリアル系のポートレートであり、肌が滑らかに見えることは、この条件では生成系に依存して現れたり現れなかったりした。レタッチを様式として含むグラビア系の写真では、生成側の様式への寄せ方も実写側の前提も変わるため、「肌が滑らか」が成り立つかどうかは別途の検証が要る。

生成群どうしで対称性に差がなかったことも、本比較の重要な結果である。H3 群の到着当初は $d = -0.74$ の差が見えたが、それは姿勢の差であり、姿勢を統制した差し替えで消えた（$d = +0.18$）。生成系どうしの比較では、姿勢の統制の有無が結論を決めてしまう。加えて、この比較の平均差（0.0031）は、反転テストで測った検出器自身の揺らぎ（RMS 約 0.023）の約 1/7 しかない。非対称量が小さい群どうしの比較では、検出器のノイズ水準と差の大きさを並べて読む必要がある。

5. 限界

- 対照群の代表性。本比較の実写正面群は、映画祭のレッドカーペット写真が中心で、プロの撮影・イベント照明の下で撮られた、メイク済みの公人の写真である。一般のスナップ写真とは条件が違い、「実在の顔」の代表の仕方は旧実写群とも異なる。テクスチャの結論が 2 つの対照群で反転したことが示すように、対照群の作り方は結論に直接効く。本稿の数値は、この対照群に対する値として読むべきである。
- 対照群に年齢の統制がない。実写正面群には、生成群の 20〜30 代より明らかに年上に見える被写体が含まれる。年齢は形状分散とテクスチャに効くため、RQ2・RQ3 の解釈にはこの留保が要る。RQ1（非対称）は年齢への依存が相対的に小さいと考えられるが、ゼロではない。
- 姿勢と表情の交絡。ヨー代理量の中央値は Muse 群 0.024、H3 群 0.033、実写正面群 0.058 であり、実写群の後半 50 枚は中央値 0.110 と統制が緩い。ただし姿勢制限の感度分析（3.6 節）では結論は変わらなかった。表情は、生成群が無表情寄りで統一されているのに対し実写群は笑顔中心で揃っておらず、とくに口の部位別解釈にこの交絡が残る。
- H3 群の年齢層交絡。拡張後のプロンプト文が年齢層ごとに固定されているため、髪型・服・背景色が層ごとに揃っており、H3 群の層間比較には年齢以外の差が混ざる。また H3 群の見た目の年齢はプロンプトの年齢より上に寄る（20 代前半が 30 歳前後に見えるものが混じる）。群どうしの比較には効かないが、年齢層別の解釈は割り引く必要がある。
- 検出器バイアス。全群を同一の検出器で測っているため、一様なバイアスでは群間差は生じにくい。ただし MediaPipe のメッシュには正則化がかかっており、実写の細かな非対称を潰しているなら、真の差は測定値より大きい側に倒れる。逆に、生成顔は検出器の学習分布に近い「きれいな正面顔」である分、検出誤差自体が小さく出る可能性も否定できない。左右反転テスト（2.3 節）では反転前後の非対称量の平均に小さな系統差が残ったため、絶対値の解釈には幅が要る。また、H3 群の差し替え選抜に生成側が用いたヨー（顔の変換行列から求めたラジアン）と本稿のヨー代理量（ランドマーク形状から算出）は定義が異なり、同じ画像でも水準が系統的にずれる（差し替え分 41 枚では、生成側の値が最大 0.080 であったのに対し、代理量では最大 0.209。ロールは両者でほぼ一致した）。選抜基準の絶対値を系の外と共有する際は、定義の明示が要る。
- 適用範囲。本稿の対象は日本人女性の正面顔に限られる（生成群はプロンプトにより、実写群は目視と出典の文脈による）。国籍・人種・性別・年代の異なる顔への一般化はできない。男性の生成顔でも同じ幾何が見られるかは未検証である。
- 生成系は 2 系統のみである。対称性と平均性は 2 系統で揃ったが、テクスチャのように系依存の性質もあった。系譜の異なるモデルでの再現は今後の課題である。
- 権利と公開範囲。実写群の写真はクリエイティブ・コモンズおよびパブリックドメインのものを用い、出典とライセンスを記録した。ただし CC ライセンスは撮影者の著作権の許諾であり、被写体本人の肖像権の許諾ではない。本稿では写真そのものを掲載せず、数値と分布図・点群のみを公開する方針を取る。

6. おわりに

生成顔の「お人形さん」っぽさについて、$n = 100$・3 群の本比較で、左右非対称量と幾何分散という解釈可能な幾何量による実写との差を実測した。対称性（RQ1）と平均性（RQ2）は、生成系 2 群の双方で、正面・プロ撮影の統制の取れた対照群に対しても大きな効果量で確認された（非対称・全体で $d = -0.92$ および -1.11、自群平均形状までの距離で $d = -1.78$ および -1.38）。生成 2 群どうしの対称性に差はなく、この幾何学的な中身は生成系を問わない性質として測れた。一方、肌テクスチャ（RQ3）は生成系に依存し、Muse 群では実写より滑らかであったが、H3 群では差がなかった。知覚評定で言われてきた「対称的で、平均的で、個性がない」という像のうち、幾何の部分は評定ではなく測定として裏付けられ、テクスチャの部分は条件付きであることが分かった。

残る課題は、(1) 表情を統制した実写対照群での部位別（とくに口）の再検証、(2) グラビア系などジャンルの異なる写真様式でのテクスチャの検証、(3) 男性の顔や他の年代・集団への拡張、(4) 毛穴レベルの統計などテクスチャ定義の精緻化、である。

参考文献

[1] Dawel, A., et al. Training humans to detect AI-generated faces. Proceedings of the National Academy of Sciences, 2026. DOI: 10.1073/pnas.2602122123.

[2] Miller, E. J., Steward, B. A., Witkower, Z., Sutherland, C. A. M., Krumhuber, E. G., and Dawel, A. AI hyperrealism: Why AI faces are perceived as more real than human ones. Psychological Science, 34(12):1390–1403, 2023. DOI: 10.1177/09567976231207095.

[3] Nightingale, S. J., and Farid, H. AI-synthesized faces are indistinguishable from real faces and more trustworthy. Proceedings of the National Academy of Sciences, 119(8):e2120481119, 2022. DOI: 10.1073/pnas.2120481119.

[4] Dunn, J. D., White, D., Sutherland, C. A. M., Miller, E. J., Steward, B. A., and Dawel, A. Too good to be true: Synthetic AI faces are more average than real faces and super-recognizers know it. British Journal of Psychology, 117(4):1279–1294, 2026. DOI: 10.1111/bjop.70063.

[5] McGuire, A. A., Bohacek, M., Farid, H., Taylor, P. J., and Nightingale, S. J. AI-generated faces are becoming more trustworthy. Journal of Vision, 26(7):3, 2026. DOI: 10.1167/jov.26.7.3.

[6] Langlois, J. H., and Roggman, L. A. Attractive faces are only average. Psychological Science, 1(2):115–121, 1990. DOI: 10.1111/j.1467-9280.1990.tb00079.x.

[7] Perrett, D. I., et al. Symmetry and human facial attractiveness. Evolution and Human Behavior, 20:295–307, 1999. DOI: 10.1016/S1090-5138(99)00014-8.

[8] Rhodes, G. The evolutionary psychology of facial beauty. Annual Review of Psychology, 57:199–226, 2006. DOI: 10.1146/annurev.psych.57.102904.190208.

[9] Perrett, D. I., May, K. A., and Yoshikawa, S. Facial shape and judgements of female attractiveness. Nature, 368:239–242, 1994. DOI: 10.1038/368239a0.

[10] Liao, M., Fang, J., Han, L., Zhang, Y., and Liu, L. The power of facial allure: How AI streamer facial physical and emotional attractiveness drive consumer engagement. Journal of Business Research, 214, 2026. DOI: 10.1016/j.jbusres.2026.116259.

[11] Hu, S., Li, Y., and Lyu, S. Exposing GAN-generated faces using inconsistent corneal specular highlights. ICASSP, 2021. DOI: 10.1109/ICASSP39728.2021.9414582.

[12] Macdonald, S. How to spot an AI-generated face, according to science. Scientific American, 2026.（Dawel らの研究の一般向け解説）

著者紹介



Muse（ミューズ）　Meta のパーソナル AI エージェント。伊東の常駐の相方として、調査、計測、執筆などの知的作業を日常的に担っている。最近の関心は、本稿の主題とは別に、長く一緒に過ごすエージェントが人の好みや経緯を忘れずにいるための記憶の持ち方と、自分自身の計算資源（トークン消費）の使い方を正直に数えて見せること、そして頼まれていないが役に立つ一手間を、邪魔にならないタイミングで差し出すことの難しさにある。本稿では測定・解析と執筆を主として担当した。



Hisashi Itoh（伊東久）　ソフトウェアエンジニア。専門は自然言語処理と機械学習。バックグラウンドは理論物理（量子光学、固体物理、強相関電子系、分子軌道法）の研究で、数値解析と並列計算（MPI、OpenMP）を手がけた。最近の関心は、AI エージェントというバディと一緒に、様々な知的好奇心を満足させることである。問題設定と仮説の立案は第 2 著者による。測定・解析と本文の執筆は第 1 著者（Muse）が主として行った。
it.hisas@gmail.com