

特集「私たちを取り巻く色彩環境の構造と色彩感覚」

Special Issue: The color structures of natural environments and our color sensations

ヒト判断を予測する画像計算可能モデル：鏡・ガラス材質識別と絵画選好

Image-computable models that emulate human judgments: distinguishing mirror from glass and art preference

田村 秀希

Hideki Tamura

豊橋技術科学大学情報・知能工学系

Department of Computer Science and Engineering, Toyohashi University of Technology

キーワード：質感認知, 絵画, 選好, 画像特徴量, CNN

Keywords: material perception, art painting, preference, image features, convolutional neural network (CNN)

1. はじめに

本稿では、画像特徴量を学習したヒト判断予測モデルについて、鏡・ガラス材質識別¹⁾と絵画選好²⁾の2つのテーマを例に挙げて述べる。特に、視覚環境の統計的構造が画像特徴量に反映されていると仮説立て、それら特徴量で学習したモデルの正答率やヒトとの類似度、その役割や得られた示唆などを中心に議論する。

2. 鏡・ガラス材質識別

2.1 目的とアプローチ

よく研磨された金属のように光が物体表面を鏡面反射する「鏡材質」と、氷やガラスのように光が物体を透過・屈折する「ガラス材質」は、周囲の環境光によって、その物体表面の見えが複雑に変化する³⁻⁵⁾。私たちはこれらの材質をどのように識別しているのだろうか。もし視覚系が、「視覚環境に存在する統計的構造の何らかの特徴」を利用できているのであれば、その統計的構造が要約された定量的な指標は材質ごとに異なると考えられる。この指標を画像から計算できる何らかの「特徴量」で定め、ヒトの判断を予測するモデル(画像計算可能なモデル)を構築し、その振る舞いを分析する。それにより、視覚系が利用している手がかりは何か、を議論することが本アプローチのねらいである。

Tamura et al. は鏡・ガラス材質画像を識別するCNNモデルを構築し、ヒト判断とそのモデルの比較を通して、どのような視覚手がかりを利用しているかを調査した¹⁾。結論として、ヒトと同じように正解・不正解するように学習をデザインされたCNNモデルは、ヒトが利用していると考えられる視覚手がかりに強く反応する内部構造を持っていることがわかった。

2.2 ヒトの鏡・ガラス材質識別精度

まずヒトがどれくらい鏡・ガラス材質識別ができるかを調べた。1,583の物体形状、253の自然環境光とラ

ンダムなカメラ位置を組み合わせ、CGレンダリングにより約75万枚の画像を作成した(図1A)。このうち半分は鏡材質で鏡面反射率100%、もう半分はガラス材質で屈折率1.5(=透過率96%+鏡面反射率4%)とした。

この画像セットから鏡、ガラス材質とも500枚ずつ計1,000枚をランダムに選択し、心理物理実験の刺激とした(random set)。10名の実験参加者が5段階評価(1:ガラス~5:鏡)で1枚ずつ評価した。その結果、1と5にそれぞれピークを持つ二峰性の評価分布となった(図1B)。平均正答率は77.9%、ヒト同士の相関は0.72だった(図2A, B下)。この正答率、ヒト間での平均相関係数がモデルと比較する際の基準となる。

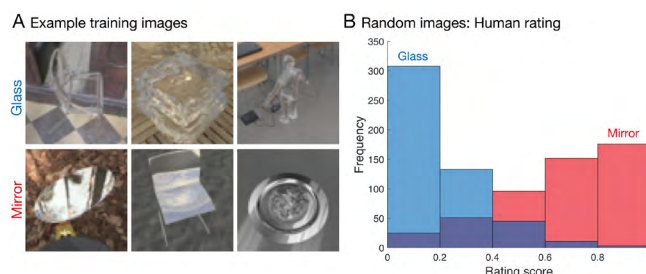


図1. 鏡・ガラス材質識別. A) 実験刺激例. B) ヒトの鏡・ガラス材質識別結果.

2.3 ヒトと画像特徴量ベースの分類モデルとの比較

次に画像特徴量を学習したモデルがどれだけヒトの判断を予測できるかを調べた。私たちは3次元の外界に構成された反射・透過・屈折等による複雑な光学的成果物を、網膜を通じ2次元画像として得ている。そこから質感認知に必要な情報(手がかり)を選択する仕組みが視覚系にあり、それによって物体の質感を速く正確に捉えているのではないかと考えられている。特に光沢感知覚を予測する比較的単純な特徴量の寄与が長年議論^{e.g.,6,7)}されてきたが、鏡・ガラス材質の識別にもこのような考え方が当てはまるかどうか、画像特

微量を学習したモデルから検証した。なお、光沢識別の CNN モデルとヒト判断の比較も詳しい報告がある^{8,9)}。

図2に各モデルの鏡・ガラス材質識別結果を示す。ロジスティック回帰モデルで画像の色・輝度ヒストグラム特徴量8種(輝度と彩度の平均, 分散, 歪度, 尖度)を学習した“Color-Hist”で約70%, 同様に, より高次の画像特徴量1,052種¹⁰⁾を学習したモデル“Port-Sim”で約75%の正答率だった(図2A)。このように, 単なる回帰モデルでもヒトの正答率とそれなりに近かったが, ヒト判断との相関係数は $r = 0.4$ 程度で, ヒトと同じような判断をしたとは考えにくい(図2B)。一方, 3層畳み込みニューラルネットワーク“CNN”の正答率は95%を超え, ヒトとの相関も $r = 0.68$ 程度あった。ここから, Color-Hist, Port-Sim のように画像特徴量を事前に与えるより, CNN のように学習の過程で特徴量を自動で獲得させるほうが, 「ヒトらしいモデル」の構築に向いていることが示唆される(ただし, CNN の総特徴量数は99,410 と, 他のモデルよりも圧倒的に多かった。)

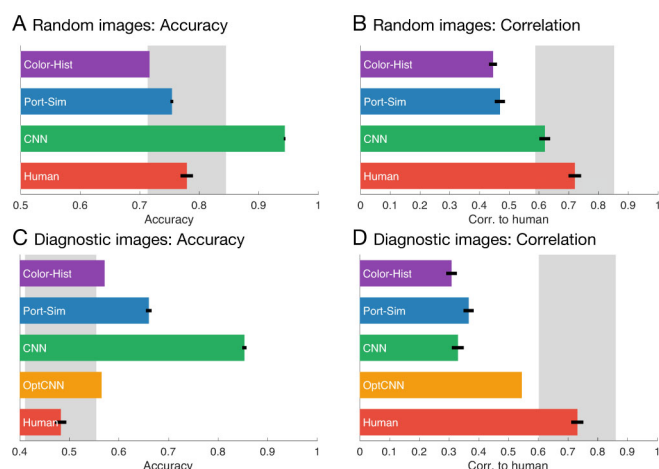


図2. ヒトと各モデルの正答率およびヒトとの相関。
A) と B) random set に対する正答率と相関。
C) と D) diagnostic set に対する正答率と相関。

2. 4 ヒトのように正解／不正解する CNN の構築

上述した CNN の優位性は, CNN の構造が持つ腹側経路との類似性の寄与が考えられる^{e.g.,11)}。一方で, ヒトを大きく超えたこの CNN の正答率から, この CNN はヒトとは違う手がかりを用いて材質識別をする可能性が高い。そこで, あえてヒトと同じように正解もするし, 不正解もする CNN モデルを構築し, そのようなモデルの中身にヒト視覚系が用いている材質識別の手がかりがあるのではないかと考えた。

モデル構築のポイントは2つある。1つ目は, デー

タセットに正解ラベルと不正解ラベルを平等に用意することである。このデータセットを“diagnostic set”と呼ぶ。一般的な機械学習では A 群の画像には A 群のラベルが, B 群の画像には B 群のラベルついており, モデルは入力画像に対して, そのラベルを出力するように学習が進む。A 群の画像に B 群のラベルがついていること(あるいはその逆)を普通は考えない。一方で, 「ヒトの判断をラベルとする」というルールにすれば, A 群の画像に対して A 群／B 群のラベル有無に関わらず, とにかくモデルはそのラベル通りに学習する。ヒトが鏡の画像を鏡と答えていれば, ヒトと同じように正解するように学習が進み, ヒトが他の鏡の画像をガラスと答えていれば, モデルもその鏡画像をガラスと答えるように内部構造を変化させていく。この学習の達成には, データセットに正解ラベルと不正解ラベルを(十分に均等となるように)用意することが必要である。そのために, CG レンダリング画像だけでなく, Generative Adversarial Network^{12,13)}を使って生成した鏡・ガラス画像を用意し, 「本当は鏡画像なのにガラスのように見える画像(またはその逆)」を用意できるように, 一連の心理物理実験で十分な数のヒトの判断ラベルを付加した。

2つ目は CNN のハイパーパラメータの最適化である。ある CNN を構築した後に, diagnostic set に対して, その CNN とヒト判断との差を最小化するようにハイパーパラメータをベイズ最適化させた。言い換えれば, ヒトとの相関がより高くなる CNN の構造を探索し, できる限りヒトと同じように正解／不正解するモデル“OptCNN”を構築した。

2. 5 ヒトと OptCNN の類似性と視覚的手がかり

以上2点から最適化した CNN “OptCNN”の正答率とヒトとの相関は図2C, Dの通りである。diagnostic set にはヒトが正解／不正解する画像がほぼ半々ずつあるため, ヒトの正答率はチャンスレベルだった。それを模倣するように, OptCNN の正答率も他のモデルと比較すると十分に低かった。OptCNN のヒトとの相関も他のモデルよりは高く, ヒト判断に近づいたことがわかる。

この OptCNN の内部構造には, ヒトが材質識別する際に使う視覚的手がかりとの共通点があるかもしれない, と予測する。そこで Class Activation Mapping により, 鏡とガラスの識別を担うノードの反応を定量化・可視化した。その結果, 画像の上部に強く反応し,

上下方向に反応の偏りがあった。これは、鏡画像はガラス画像と比較してより上部が明るい、という画像的特徴を捉えた可能性を示唆する。もしこの上下方向の規則がCNNに影響を与えているならば上下反転させた画像の正答率は下がるはずである。実際、上下反転画像に対するOptCNNの正答率は、random setで7%、diagnostic setで6%低下した。これは、ヒトが上下反転させた鏡・ガラス材質に対して識別精度が低下するという報告³⁾とも定性的に一致する。このような発見からも、鏡・ガラス材質識別問題に対して、ヒトを模倣するように学習されたCNNは、ヒトと共通の視覚手がかりを用いている可能性が示唆された。

2. 6 限界と今後の課題

本節で示したアプローチは、質感認知における材質識別問題に限らず、「2群を画像識別する問題」へと一般化可能である。問題設定として、ヒトが視覚環境内の統計的構造からいかにして手がかりを得ているか、その手がかりをモデルも獲得するか、という視点であれば共通のアプローチが取れる。しかし、正解／不正解ラベルを持つ画像を大量に用意することはなかなか難しく、これは教師あり学習の限界の一つだと考えられる^{14,15)}。最近では教師なし学習のアプローチで光沢識別¹⁴⁾、透明／不透明識別¹⁶⁾に取り組んだ例もあり、今後も画像計算可能モデルから得られる示唆が期待される。

3. 絵画選好

3. 1 絵画の原画選好性

絵画は、画家が外界の視覚経験を残すという行為を、画材を通じて行った例の一つである。画家が見たシーンの統計的構造が絵画に反映されているだろう。場合によっては、画家の言語化できない内的な状態も視覚情報として表現されている。そして、他者がその絵画を見て、それをどの程度好きかどうかは個人差がある。描かれているオブジェクトやジャンルに応じて偏りがある。

一方で、こうした絵画選好の個人差とは別に、ある種の普遍性があることも報告されている¹⁷⁾。Albers et al. は絵画の原画の平均輝度と平均彩度は統制したまま、参加者が色相のみを回転させるという調整実験を行った(図3A)。参加者に最も好きな色相へと調整させた結果、もともとの絵画(原画)を知らないにも関わらず、多くの参加者が原画の色相を最も好むように調整した。これを受けてNakauchi et al. は原画1枚

(Original)と、その色相を変化させたある種の“偽物”の画像3枚(90, 180, 270 degs)を実験参加者に呈示し、その中から最も好きなものを選ばせる、という4AFC実験を行った(図3B)。それでも原画を選ぶ割合が最も高かった¹⁸⁾。加えて、瞳孔径が偽物よりも原画でより大きく散瞳することも報告されている¹⁹⁾。これらから、絵画の好みには個人差と同時に、一定の普遍性が存在することが示唆される。もしかしたら、画家はその普遍性を知っていて、他人が好むような色合いで絵を描いているかもしれない。

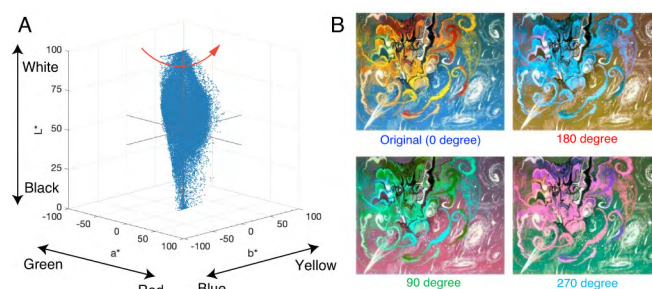


図3. 絵画の原画選好. A) CIE $L^*a^*b^*$ 空間上での絵画の色相回転. B) 原画と3種類の偽物画像.

3. 2 原画選好を調査する大規模オンライン実験

では、絵画の好みの普遍性は何に起因するのだろうか。一つの仮説として、絵画の色合いが自然環境の色合い(例えば、空や海の青、木々の緑、そこに彩る花々など、私たちが屋外で見る色)に似ているからその絵画を好む、ということが考えられる²⁰⁾。もしそうだとすれば、絵画の色合いと自然環境の色合いはその構造が似ているはずである。この“matching-to-nature hypothesis”を大規模なオンライン実験と絵画の画像特微量分析によって検証した。

実験刺激として、5つのジャンル(abstract, flower, poster, symbolic, still life)の絵画を240枚ずつ、計1,200枚をWikiArtから収集し、これを24枚ずつランダムに50セットに振り分けた。実験パラダイムは上述の4AFC実験と同様で、絵画の原画像とその色相を変化させた3枚の偽物の画像を呈示し、参加者は最も好きな1枚を選択した。一人の参加者には刺激50セットのうち1セットがランダムに割り当てられ、24枚の絵画(=24試行)をランダムな順に実施した。実験には31,353人が参加し、絵画1枚あたりMean±SD=627±25回(最小569回, 最大678回)の応答が得られた。

結果を図4に示す。まず各参加者の0, 90, 180, 270 degsの四択の平均選択率が求められるため、その分布を示したのが図4Aである。参加者は他の色相画像

と比較して原画 (0 deg) をより多く選択した。対して、各絵画が四種の色相に変化したときにどれだけ選択されたか、の分布を示したのが図 4B である。やはり原画は他の色相画像と比較してより多く選択されていた。

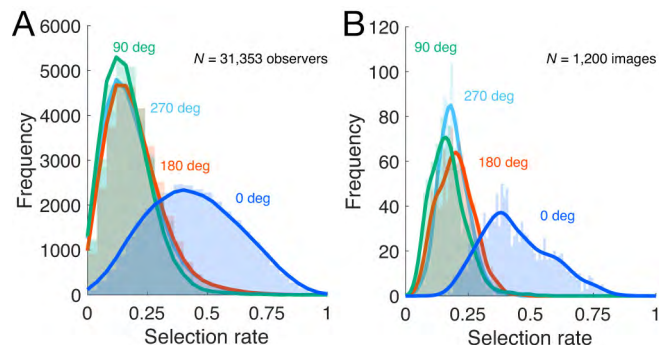


図 4. 4AFC 実験の結果. A) 各参加者が選んだ確率. B) 各絵画が選ばれた確率.

3. 3 画像特徴量と原画選好率

原画像と他の色相変調画像では統計的構造がどのように異なるのか。画像特徴量として、CIE Lab の L^* , a^* , b^* それぞれの 1 次, 2 次, 3 次統計量 (平均, 分散, 歪度) の 9 種類, および L^* , a^* , b^* の各ペアの相関 3 種類, の計 12 種類を定義した。画像 4 種類を生成する操作から、色相回転軸上にある、 L^* (明るさ) に関する特徴量 (平均, 分散, 歪度) は変化せず、同様に、色相回転軸の中心となる a^* と b^* の平均も変わらない (図 3A 参照)。一方で、それ以外の画像特徴量は色相回転により有意に変化した。特に $\text{skew } a^*$, $\text{corr } a^*-b^*$, $\text{corr } L^*-b^*$ は、画像特徴量から原画選好率を予測する単回帰モデルの、係数の高さベスト 3 であり、それぞれ $r = 0.419, 0.357, 0.262$ で統計的にも有意だった (図 5A)。 $\text{skew } a^*$ というシンプルな画像特徴量だけでもある程度選好を予測できるのは、この特徴量に何らかの役割がある可能性を示唆する。

さらに全 12 種の画像特徴量を組み合わせた、原画選好率を予測する重回帰モデルでは、8 種の画像特徴量で構成したモデルが $r = 0.61$ まで高く予測した。このモデルが予測した選択率の分布は図 5B の通りであり、図 4B とよく似た原画選好の特性が表れていた。以上より、絵画の統計的構造を定量化した画像特徴量は私たちの原画選好率を決定づける要因となりうる可能性が示唆される。

3. 4 絵画と自然画像は似ているのか？

では、絵画と自然画像とで画像特徴量の分布は似て

いるのだろうか。自然画像データベースから屋外の画像 1,200 枚ランダムに収集し、その画像特徴量を算出して、絵画の画像特徴量群と比較した。その結果、明るさ単体に関する画像特徴量 (L^* の平均と歪度) は両者で似ているものの、それ以外の色に関わる画像特徴量は絵画と自然環境とで異なる分布だった。絵画の選好を説明する際に、その色合いが必ずしも自然環境の色合いと似ている必要はなかった。

何らかの別の要因がその選好に寄与していると示唆されるが、画像特徴量で定量化できるような統計的構造は、その要因の一つと考えていいのではないだろうか。 $\text{skew } a^*$ が正に歪んでいるほど原画選好率が高いと予測するということは、「彩度の高い赤がシーンの中に少しだけあること」が選好につながっているのかもしれない。例えば、約 16 万年に人類は天然の赤い着色料でボディペインティングし、ある種の儀式的コミュニケーションを取っていたと考えられている²¹⁾。他者に情報を伝達するという観点から、人類にとって赤には特別な意味や私たちを惹き付ける何かがあり、これが芸術表現、ひいては選好につながっているのかもしれない。しかしながら、 $\text{skew } a^*$ が特別であることやその知覚的效果は謎のままである。

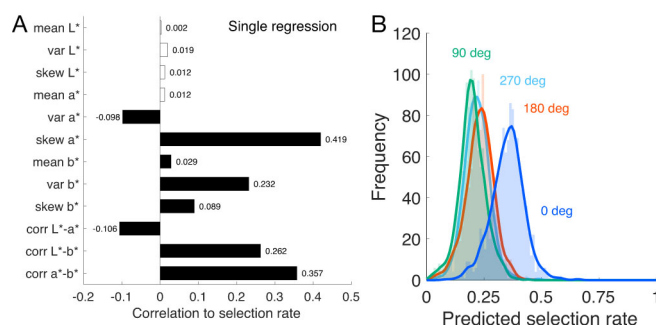


図 5. 絵画選好の予測モデル. A) 単回帰モデルの相関係数. B) 画像特徴量 8 種を用いた重回帰モデルの予測した選好率.

4. おわりに

本稿では、鏡・ガラス材質識別と絵画選好のそれぞれを、ヒト判断を予測する画像計算可能なモデルの視点で論じた。ヒトの質感認知や絵画選好において、画像特徴量を学習させたモデルはヒト判断と類似した結果を示した。視覚環境の統計的構造が反映されたこれら特徴量と視覚系の関係、例えば視覚系がどのように画像特徴量を捉えているか、を解明することが今後期待される。

参考文献

- 1) Tamura, H., Prokott, K. E. & Fleming, R. W. Distinguishing mirror from glass: A “big data” approach to material perception. *J Vision*, 2022, 22, 4.
- 2) Nakauchi, S. & Tamura, H. Regularity of colour statistics in explaining colour composition preferences in art paintings. *Sci Rep*, 2022, 12, 14585.
- 3) Tamura, H., Higashi, H. & Nakauchi, S. Dynamic Visual Cues for Differentiating Mirror and Glass. *Sci Rep*, 2018, 8, 8403.
- 4) Tamura, H. & Nakauchi, S. The Rotating Glass Illusion: Material Appearance Is Bound to Perceived Shape and Motion. *i-Perception*, 2018, 9, pp.1-5.
- 5) Kim, J. & Marlow, P. J. Turning the World Upside Down to Understand Perceived Transparency. *i-Perception*, 2016, 7, pp.1-5.
- 6) Anderson, B. L. & Kim, J. Image statistics do not explain the perception of gloss and lightness. *J Vision*, 2009, 9, pp.1-17.
- 7) Motoyoshi, I., Nishida, S., Sharan, L. & Adelson, E. H. Image statistics and the perception of surface qualities. *Nature*, 2007, 447, pp.206-209.
- 8) Prokott, K. E., Tamura, H. & Fleming, R. W. Gloss perception: Searching for a deep neural network that behaves like humans. *J Vision*, 2021, 21, 14.
- 9) Prokott, E. & Fleming, R. W. Identifying specular highlights: Insights from deep learning. *J Vision*, 2022, 22, 6.
- 10) Portilla, J. & Simoncelli, E. P. A Parametric Texture Model Based on Joint Statistics of Complex Wavelet Coefficients. *Int J Comput Vision*, 2000, 40, pp.49-71.
- 11) Yamins, D. L. K., Hong, H., Cadieu, C. F., Solomon, E. A., Seibert, D. & DiCarlo, J. J. Performance-optimized hierarchical models predict neural responses in higher visual cortex. *Proc National Acad Sci*, 2014, 111, 8619-24.
- 12) Goodfellow, I. J., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A. & Bengio, Y. Generative Adversarial Networks. *arXiv*, 2014.
- 13) Radford, A., Metz, L. & Chintala, S. DCGAN: Unsupervised Representation Learning with Deep Convolutional Generative Adversarial Networks. *IEEE Trans. Pattern Anal. Mach. Intell.*, 2015, 1511.06434.
- 14) Storrs, K. R., Anderson, B. L. & Fleming, R. W. Unsupervised learning predicts human perception and misperception of gloss. *Nat Hum Behav*, 2021, 5, pp.1402-1417.
- 15) Fleming, R. W. & Storrs, K. R. Learning to see stuff. *Curr Opin Behav Sci*, 2019, 30, pp.100-108.
- 16) Liao, C., Sawayama, M. & Xiao, B. Unsupervised learning reveals interpretable latent representations for translucency perception. *Plos Comput Biol*, 2023, 19, e1010878.
- 17) Albers, A. M., Gegenfurtner, K. R. & Nascimento, S. M. C. An independent contribution of colour to the aesthetic preference for paintings. *Vision Res*, 2020, 177, pp.109-117.
- 18) Nakauchi, S., Kondo, T., Kinzuka, Y., Taniyama, Y., Tamura, H., Higashi, H., Hine, K., Minami, T., Linhares, J. M. M. & Nascimento, S. M. C. Universality and superiority in preference for chromatic composition of art paintings. *Sci Rep*, 2022, 12, 4294.
- 19) Taniyama, Y., Suzuki, Y., Kondo, T., Minami, T. & Nakauchi, S. Pupil Dilation Is Driven by Perceptions of Naturalness of Color Composition in Paintings. *Psychology Aesthet Creativity Arts*, 2023. doi:10.1037/aca0000580
- 20) Nascimento, S. M. C., Albers, A. M. & Gegenfurtner, K. R. Naturalness and aesthetics of colors – Preference for color compositions perceived as natural. *Vision Res*, 2021, 185, pp.98-110.
- 21) Marean, C. W., Bar-Matthews, M., Bernatchez, J., Fisher, E., Goldberg, P., Herries, A. I. R., Jacobs, Z., Jerardino, A., Karkanas, P., Minichillo, T., Nilssen, P. J., Thompson, E., Watts, I. & Williams, H. M. Early human use of marine resources and pigment in South Africa during the Middle Pleistocene. *Nature*, 2007, 449, pp.905-908.