

# 画像分類逆相関法で生成された顔のイメージが表すもの —ディープラーニングに基づく顔認識モデルからの検討—

上田 祥行

京都大学人と社会の未来研究院

In the reverse correlation image classification technique for faces, classification images (CIs) are constructed by collecting the superimposing polar inversion noise onto a base face image. CIs reflect the features participants used in their judgments in a bottom-up manner and are considered to visualize their mental representations. However, whether the features of the generated CIs correspond to the actual face that participants viewed is typically assessed through the subjective judgment of a third person. In this study, I applied a deep learning-based face recognition model (Serengil & Özpınar, 2024) to the generated CIs to evaluate their similarities to the original face. The results showed that the similarity between the CIs and the original face became stable after more than 200 trials of the classification task. Moreover, while the CIs captured characteristics of the original face more than other faces, the face recognition model could not accurately identify which face was viewed. These findings suggest that the face represented by the reverse correlation image classification is not an exact replica of the target face, but is perceived as more similar to the face of the target than to other faces.

Keywords: image classification reverse correlation, face, mental representation, face recognition model, deep learning.

## 問題・目的

顔の記憶は様々な要因により比較的簡単に歪められてしまう。人々が持つ顔の記憶がどのようなものかを検討するため、これまでの研究ではしばしば再認課題が用いられてきた。この課題では、ターゲットの顔を記憶し、後に呈示される顔が記憶した顔と同じかどうかを判断することが求められる。テスト刺激の特定の領域（例えば、目の周辺）を変化させることでその領域が記憶に与える影響を調べることができるが、どの領域をどのように変化させるのかは研究者の仮説に依存しており、実際に人々がどのような記憶を保持しているかとは必ずしも一致しないという問題があった。

近年提案されている画像分類逆相関法 (reverse correlation image classification technique: RCIC) はこの問題を克服し、人々が持つ顔の表象を直接描き出すことができる方法として注目されている (Mangini & Biederman, 2004; Todorov et al., 2011)。RCIC では、ベースとなる顔画像 (通常は複数の顔の平均顔) に極性を反転させて作成した2種類のノイズを付加した刺激ペアを呈示し、参加者にどちらが記憶した顔に近いかを選択させる。選択された刺激に付加されたノイズを集約して作成された分類画像 (classification image: CI) は、参加者が判断に用いた部分がボトムアップに強調され、参加者が保持していた顔の表象を可視化していると考えられている。

しかし、このようにして作成された CI が、記憶した顔の特徴と一致しているかどうかは、現時点では第三者の主観判断によって評価されており、客観的な指標が存在しない。そこで本研究では、人間の顔認識能力に迫る精度を達成したディープラーニングベースの顔認識モデルを用いて、作成された CI が実際に記憶した顔とどの程度類似しているのかを評価した。このようなモデルを活用する利点は、2つの顔の類似度を数値化できる点にある。また、RCIC の手法を用いた実験は多くの試行数が必要となるため、顔認識モデル

に大量の CI を評価させることで、安定した CI を得るために必要な試行数についても検討した。

## 方法

### 顔認識モデル

Python向けの顔認識フレームワークであるDeepFace (Serengil & Özpınar, 2020) を使用した。顔認識モデルに Facenet512、顔検出モデルに retinaface を採用し、距離尺度としてコサイン類似度を設定した。この組み合わせは、Labelled Faces in the Wild (LFW) データセットによるテストにおいて、ヒトの顔認識能力をわずかに上回る精度 (人間の正答率: 97.5%、顔認識モデルの正答率: 98.4%) を示した (Serengil & Özpınar, 2024)。

### 刺激

上田ら (2023) の顔記憶実験における RCIC のデータを分析対象として使用した。この実験の RCIC の手続きは標準的なものであり、以下にその概要を示す。刺激の作成には、Rのrcicrツールボックス (Dotsh, 2022) が用いられた。

実験参加者は米国在住の成人180名であり、Chicago Face Database (Ma et al., 2002) に含まれる4名 (男女2名ずつ) の顔のうち、いずれか1名の顔を1分間記憶した。一定の遅延の後、男性の顔を記憶した参加者には2名の男性の平均顔に極性反転ノイズを付加した刺激対が、女性の顔を記憶した参加者には2名の女性の平均顔に極性反転ノイズを付加した刺激対が呈示された。参加者は、どちらが記憶した顔に似ているかを640試行にわたって判断した。1人の参加者に対し、各試行で付加されたノイズはすべて異なっていたが、同じ性別の顔を記憶した参加者間では、同じノイズが順不同で用いられた。

参加者が選択した刺激に付加されたノイズを集約し CI を作成した。CI の作成にあたっては、各参加者について1~10 試行目の結果を用いた CI、1~20 試行目の結果を用いた CI、1~30 試行目の結果を用いた CI、と10 試行ずつ使用する試行数を増やし、最終的には

1~640 試行目のすべての結果を用いた CI まで、64 種類の CI を作成した。また、同じ顔を記憶した参加者の結果を集約し、同様に 10 試行ずつ使用する試行数を増やして、記憶された顔ごとに 64 種類の CI を作成した。この手続きにより、全部で 11,776 枚の CI (180 人 × 64 種類 + 4 つの顔 × 64 種類) が顔認識モデルによって評価された。記憶された顔の例と、これらの顔を記憶した参加者のすべての試行を集約して作成した CI を Figure 1 に示す。

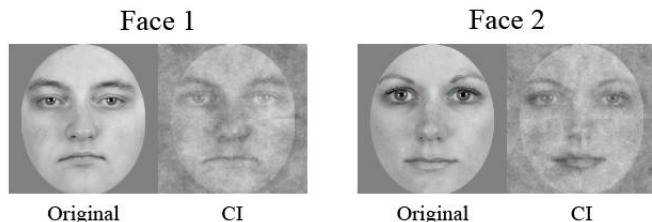


Figure 1. Original faces and generated classification images (CIs).

## 結果

抄録の紙面の都合上、ここでは女性顔を記憶したときの CI に対する評価結果のみを示す。

顔認識モデルによって、記憶された顔と同一人物であると判定された CI は極めて少なかった。640 試行すべての結果を使用した場合でも、元画像と同一人物であると判定された割合は、Face 1 では 0% (0/46)、Face 2 では 4.7% (2/43) にとどまった。

試行数の変化に伴う類似度の変化を Figure 2 に示す。この図では、Face 1 との類似度を示している。試行数が増えるにつれて、Face 1 を記憶した参加者の CI は徐々に Face 1 と近づく傾向が見られた。

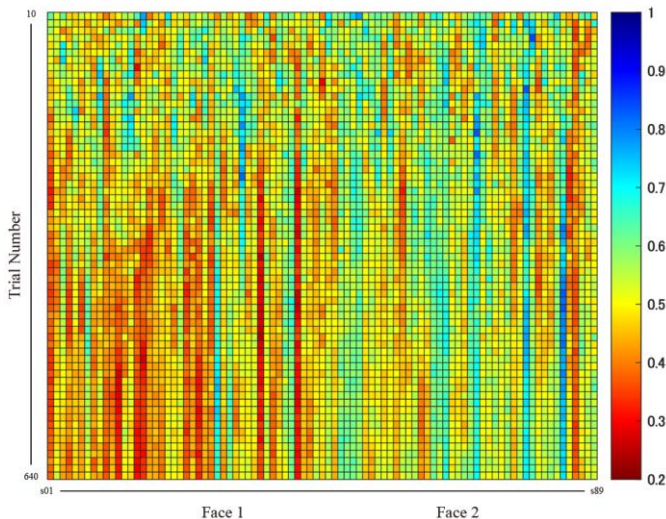


Figure 2. Distances between the original face (Face 1) and CI. Columns represent participants, while rows represent trial numbers.

1枚の CI に対して、記憶した顔との類似度および記憶していない顔との類似度を計算し、その差分を示した結果を Figure 3 に示す。この図では、Face 2 に比べて Face 1 の類似度が高いほど赤く、Face 1 に比べて Face 2 の類似度が高いほど青く表現されている。この結果から、Face 1 を記憶して作られた CI は Face 2 よ

りも Face 1 に類似しており、Face 2 を記憶して作られた CI は Face 1 よりも Face 2 に類似していると評価されていることが示された。また、同じ参加者のデータ (同じ列に記載されたデータ) でも、試行数が少ない段階では赤と青が入れ替わっているのに対し、200 試行を超えたあたりからは色の入れ替わりが減少し、安定的に一方の顔に類似していると評価されることが示された。すべての参加者の試行を集約して作成された CI は、各参加者の試行数が少ない段階 (40 試行程度) から安定することが明らかになった。

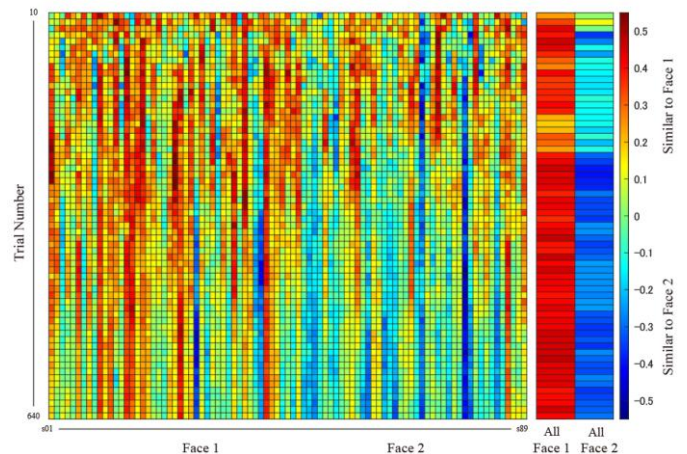


Figure 3. The degree of similarity across the original faces and CI. The rightmost columns show collective data. Red color indicates greater similar to Face 1, while blue color indicates greater similar to Face 2.

## 考察

本研究では、RCICの手法を用いて作成された CI の妥当性を、ディープラーニングに基づく顔認識モデルを適用して検討した。その結果、CI は記憶した顔そのものと同一人物だとは判断されないものの、他の顔と比較すると、記憶した顔により類似していると判断された。また、200 試行程度の分類課題を行うことで、この傾向は安定することも明らかになった。本研究の結果は、RCICによって作成された CI は「記憶した顔そのものではないが、他の顔と比べると記憶した顔により類似している」ものであることを示唆している。

本研究で用いた顔認識モデルは、ヒトとほぼ同じ精度の顔認識成績を示すものであるが、ヒトが CI に対して同様の判断をするかは未だ明らかではない。今後、本研究の結果がヒトによる評価で再現可能かどうかを検討する必要がある。また、本研究では限られたデータセットを用いたため、データセットを拡張し、本研究結果の一般化の可能性についても検討する。

## 引用文献

Serengil, S. & Özpınar, A. (2020). In *2020 Innovations in Intelligent Systems and Applications Conference (ASYU)*, 23-27.

Serengil, S. & Özpınar, A. (2024). *Bilişim Teknolojileri Dergisi*, 17(2), 95-107.

上田ら. (2023). 日本認知心理学会第21回大会, O-D03.