



THE VIRTUAL REALITY SOCIETY OF JAPAN

# 畳み込みオートエンコーダを用いた並置混色錯視の知覚再現評価

Evaluation for Machine Perception of Color Mixture Illusion Using Convolutional Autoencoder

久保田祐貴<sup>1)</sup>, 稲見昌彦<sup>2)</sup>

Yuki KUBOTA<sup>1)</sup>, Masahiko INAMI<sup>2)</sup>

1) 東京大学 大学院情報理工学系研究科 (〒 113-8656 東京都文京区本郷 7-3-1, yuki\_kubota@ipc.i.u-tokyo.ac.jp)

2) 東京大学 先端科学技術研究センター (〒 153-8904 東京都目黒区駒場 4-6-1)

**概要:** 錯視は人間の多様な知覚・認知機能を解明する手法として注目されてきた。特に機械学習を用いた錯視再現は、知覚再現に基づく感覚操作の試金石となりうる。本研究では、畳み込みオートエンコーダを用いて並置混色錯視の定量評価を行った。ノイズ除去やブラー除去を学習した学習機に画像ピクセル値を変更した並置混色錯視を入力した結果、各々の学習特性に関連した錯視知覚特性を得た。また、学習の進行とともに錯視知覚が連続的に変化する様子が観察された。

**キーワード:** 錯視再現, 機械学習, 感覚・知覚操作, 心理物理実験, 知覚評価

## 1. はじめに

錯視は脳の視覚処理が表出した現象であり、人間や動物の形・色・動きなど多様な知覚・認知機能を解明する手法として古くから注目されてきた [1, 2]。また、近年の情報科学分野では、画像処理や機械学習の枠組みで視覚処理モデルを作成し、計算機上で錯視現象を再現する *kennkyuu* が行われている [3, 4]。計算機における錯視再現は、錯視・錯覚を手がかりとして感覚や知覚を計算機上に再現し、その知覚再現に基づいて感覚操作や知覚介入を行うことが可能となる点で、バーチャルリアリティ (Virtual Reality: VR) 分野における基盤的な研究としても重要である [5]。

特に、機械学習を用いた知覚再現は、予め視覚処理モデルを与える手法に比べてデータの学習によるモデルの可塑性が期待できることから研究が進められている。例えば、一人称視点で撮られた動画を深層学習した計算機が未学習の蛇の回転錯視を知覚したと報告された研究 [6] や、ブラー除去やノイズ除去を学習した畳み込みオートエンコーダ (Convolutional Autoencoder: CAE) が未学習の色や明るさに関する錯視知覚を部分的に再現した研究がある [7]。しかし、色や明るさに関する錯視において、画像パラメータにより学習済モデルの錯視再現性能がどのように変化するかは検討されていない。

本研究では、ブラー除去やノイズ除去を学習した CAE を用いて、並置混色錯視に関する錯視知覚の定量評価を試みる。並置混色錯視とは、図 1 に示すような Munker-White 錯視 [8, 9] としても知られ、2 種類の色や輝度を持つ縦縞に影響されて縦縞内に配置された図形の色や輝度が変化する現象である。図 1 の縦縞内に存在する灰色の正方形は左右で輝度が等しいが、白帯が正方形上に露出した右の画像がより明るく知覚される。色画像の場合も同様に、2 色の帯の

どちらが正方形上に露出しているかにより知覚される色や輝度変化する。Munker-White 錯視は、単純な同時対比効果で予測される知覚と逆の知覚を示す [10]。そのほか、ストライプの輝度 [11] や空間周波数 [12] の依存性について研究が報告されている。本研究では特に、輝度に変化する白黒の Munker-White 錯視を取り上げ、ブラー除去またはノイズ除去を学習した CAE が輝度をどのように推定するかについて定量的に評価する。

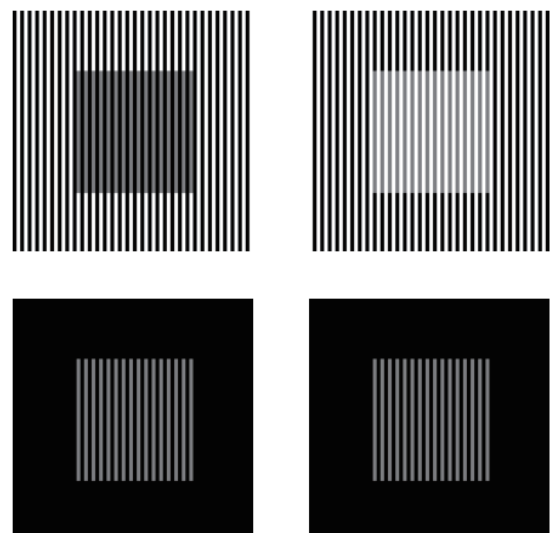


図 1: Munker-White 錯視の刺激画像例。上段左右の画像の中央部に位置する正方形は同じ輝度であるが、周囲の縦縞の影響により右側の正方形の輝度がより明るく知覚される。下段は左右それぞれの画像に対し、中央の正方形が露出している部分のみにマスクをかけたもの。

## 2. 使用した学習モデルおよび学習過程

学習モデルとして、図2に示すような隠れ層を1層のみ持つ単純な構造のCAEを採用した。128x128の8-bit RGB画像(128x128x3)を入力とし、この入力を、シグモイド関数を活性化関数に持ち、カーネルサイズが5である畳み込み層により128x128x8の隠れ層へと出力する。その後、同様の畳み込み層によって128x128x3の形で出力する。

学習データを選定する上で、最終的に並置混色錯視の平面的な刺激画像と比較することを考慮する必要がある。本研究では、自然画像を多く含むテクスチャデータセットであるDescribable Textures Dataset (DTD) [13]を採用した。画像数は5096枚であり、全てを学習データに用いた。



図2: 本研究で使用する学習モデルのネットワーク構造。

ノイズまたはブラーを付加した全ての学習データをシャッフルして128個のバッチに分割したのち、100エポックの学習を行った。ただし、損失関数の値が3エポック以上改善しない場合、学習の早期打ち切りを行った。Adamsにより最適化を行い、損失関数として入力と出力の平均二乗誤差(Mean Squared Error: MSE)を用いた。ノイズ除去学習では、 $\sigma = 16$ のガウシアンノイズを付加した画像を入力として、ノイズ付加前の画像を正解出力とした。ブラー除去学習では、 $\sigma = 2$ のガウシアンブラーを付加した画像を入力として、ブラー付加前の画像を正解出力とした。入力と出力のMSEを求め、誤差を逆伝播させてネットワーク中の重みを更新した。

なお、ノイズ除去およびブラー除去のタスクは、学習データにノイズまたはブラーを付加した画像を用意することでCAEの枠組みで実装を行うことができ、学習タスクの特性が人間の知覚特性とどのような関わりを持っているかを議論できると考えられたために採用した。また、機械学習において低次のレベルの画像処理の問題として複数の研究が取り扱っているタスクであり[7, 14, 15]、これらの先行研究で提案されたモデルとの比較が可能であると考えられる。

## 3. 錯視知覚量の定義

並置混色錯視の定量評価を行うためには、モデルの入出力に基づいて、並置混色錯視の錯視知覚を特徴付ける画像パラメータを計算する必要がある。

並置混色錯視の特徴は、黒縞と白縞の配置によって中央部の正方形の輝度知覚が変化する点にある。そのため、本

研究における並置混色錯視の錯視量を次のように計算した。まず、中央の正方形部分のうち、2種のピクセル値 $P_1$ と $P_2$ を持つ縞によって覆われていない部分のみをマスクにより取り出した。この取り出した画素値の平均を各画像における正方形の知覚輝度とした。ただし、学習にRGB画像を用いたため、CIE XYZ表色系の輝度(Y)の換算に従い、

$$I = 0.2126 \times R + 0.7152 \times G + 0.0722 \times B \quad (1)$$

で輝度値 $I$ に変換したのち平均をとった。さらに $P_1$ ,  $P_2$ の縞が正方形の裏側に隠れる画像をそれぞれ画像1, 画像2としたとき、画像1および画像2の出力画像から計算された知覚輝度をそれぞれ $I_1$ ,  $I_2$ とした。このとき、最終的な並置混色錯視の錯視量 $\Delta I$ を $\Delta I = |I_1 - I_2|$ として計算した。

同時に、入出力としてカラー画像を取っているため、R, G, B個々のピクセル値も出力される。 $\Delta I$ 以外の指標として、画像 $i$ のR, G, Bそれぞれの成分のピクセル値 $I_{iR}$ ,  $I_{iG}$ ,  $I_{iB}$ も計算した。

## 4. 評価実験の手法

学習モデルは2.で述べた隠れ層を1層持つCAE、学習タスクはブラー除去(deblur)とノイズ除去(denoise)の2条件とした。入力画像として、図1に示すような画像サイズが128px四方の8-bit RGB画像を用い、ピクセル値 $P_S$ の64px四方の中央の正方形は対角線中心が画像の中心と一致するように配置した。ピクセル値 $P_1$ と $P_2$ からなる縞は線幅 $W$ で画像の左端が $P_1$ となるように交互に配置した。

本研究では、正方形のピクセル値 $P_S$ を変更した場合の応答を検証した。線幅 $W$ は4px,  $P_1 = 255$ ,  $P_2 = 0$ とし、 $P_S$ を0から255まで256段階で変化させた。学習モデルはいずれもカラー画像に対して学習を行わせたため、 $P_1$ や $P_2$ は $(R, G, B) = (255, 255, 255)$ という標識であることに注意が必要である。評価値は、出力画像の輝度 $I_i$ および2つの画像の輝度差 $\Delta I$ , RGB各成分の輝度値 $I_{iR}$ ,  $I_{iG}$ ,  $I_{iB}$ である。ここで、 $P_1$ と $P_2$ の縞が正方形の裏側に隠れる刺激画像をそれぞれ画像1, 画像2とした。

## 5. 結果と考察

図3(a), (b)にそれぞれブラー除去およびノイズ除去を学習したCAEの出力結果を示す。各図の左側は $I_1$ (赤実線),  $I_2$ (赤点線),  $\Delta I$ (青実線)、右側は $I_{iR}$ ,  $I_{iG}$ ,  $I_{iB}$ ( $i = 1$ : 実線,  $i = 2$ : 点線)の結果を示す。いずれも横軸が入力画像のピクセル値、縦軸が計算された出力輝度である。

まず、各図の左側に示した $I_i$ および $\Delta I$ の結果についてまとめる。 $I_1$ と $I_2$ の大小関係についてみると、ブラー除去を学習したモデルは $I_1 > I_2$ を予測している一方で、ノイズ除去を学習したモデルは $I_1$ と $I_2$ の間に差が見られない。実際には、図1の上段に示すように $I_1 < I_2$ であることから、ブラー除去条件は人間の知覚とは逆の予測を示しており、ノイズ除去条件は入力画像に近い値を出力している。ブラー除去条件では $I_1 > I_2$ , すなわち周囲の輝度が暗いほど

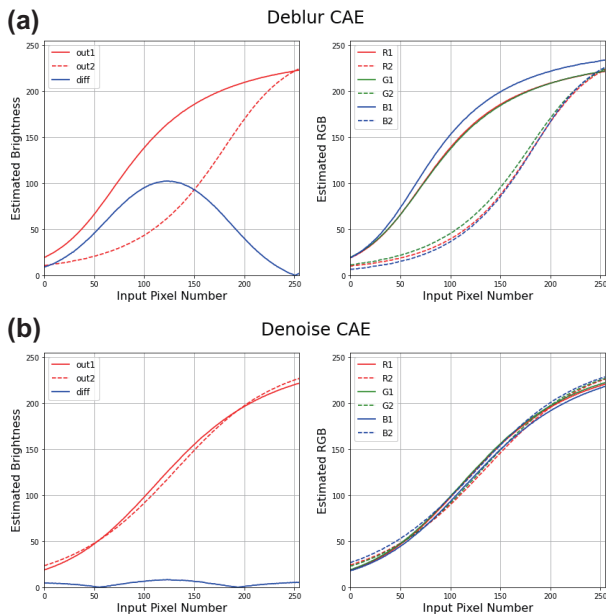


図 3: 実験の結果. 各図の左側は  $I_1$  (赤実線),  $I_2$  (赤点線)  $\Delta I$  (青実線) を示し, 右側は  $I_{iR}$ ,  $I_{iG}$ ,  $I_{iB}$  を示す ( $i = 1$ : 実線,  $i = 2$ : 点線). (a) ブラー除去を学習した CAE の出力結果. (b) ノイズ除去を学習した CAE の出力結果.

中央の正方形が明るく見えるという同時対比効果を予測しており, ブラーによって鈍った輝度を鮮明化する学習が行われたと考えられる. 一方, ノイズ除去条件では局所的に存在するガウシアンノイズのみを除去しノイズ周辺の画素を変更しない学習を進めたために, 入力画像と出力画像のピクセル値に差が見られなかったと考えられる. これらの結果は, 低レベルの視覚処理に関する学習タスクが人間の視覚特性の一部と対応することを示唆する.

次に, 各図の右側に示した  $I_{iR}$ ,  $I_{iG}$ ,  $I_{iB}$  の結果についてまとめる. ブラー除去を学習したモデルでは, 高輝度側において  $I_{1B}$  がより大きく青色, 低輝度側において  $I_{2R}$  や  $I_{2G}$  が大きく黄色みを帯びた色味が予測されている. 図 1 の下段に示すように, 実際の人間の知覚でもマスク後の画像に対して黄色みや青みを帯びた灰色の縦縞を観察することができる. 入力画像がマスク適用前の画像であることに注意する必要があるが, これらの結果は黄色みあるいは青みを帯びた灰色を知覚する現象が学習モデルにより強調されたことを示す可能性がある. 一方で, これらの色味は学習データのバイアスによって生じている可能性があるため, 学習データを変更したモデルや異なる学習パラメータを採用した際にこれらの性質に頑健性があるかどうかを調査する必要がある.

さらに, 学習モデルの学習過程と錯視知覚の変化の対応関係について述べる. 図 4 にブラー除去を学習したモデルの学習過程における錯視知覚の変化を示す. 赤から青にかけての色のグラデーションは, 赤が学習初期, 青が学習終了時とした際の学習の進行を示す. また, 図 4(a) が  $I_1$ , 図

4(b) が  $I_2$  の推定値を示す. いずれも横軸が入力画像のピクセル値, 縦軸が計算された出力輝度である.

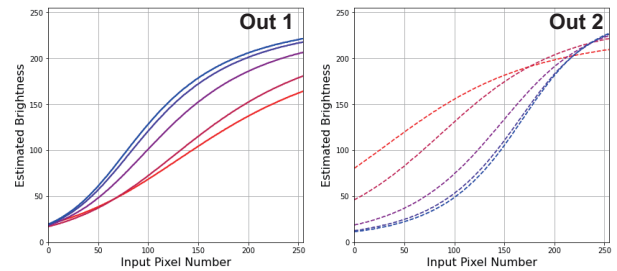


図 4: ブラー除去を学習したモデルの学習過程における錯視知覚の変化. 赤から青にかけてのグラデーションは, 赤が学習初期, 青が学習終了時として学習の進行を示す. (a) が  $I_1$ , (b) が  $I_2$  の推定値を示す.

学習過程での錯視知覚の変化を見ると,  $I_1$  は学習の進行とともに単調に増加し,  $I_2$  は学習の進行とともに単調に減少していることが見て取れる. さらに, 学習初期は  $I_1 < I_2$  であるが, 学習が進むとその差が小さくなったのち知覚輝度が逆転し,  $I_1 > I_2$  となっている. このようにブラー除去の学習が進むに従って, 予測される知覚輝度が連続的に変化している. これらの結果は, 錯視知覚を利用して学習過程や学習の進行度を評価できることを示唆する. 他の錯視知覚に対する出力も合わせて調査することで学習過程をどのように把握できるかを多角的に検討することが今後の課題となる.

## 6. 結論と今後の課題

本研究では, 畳み込みオートエンコーダ (CAE) を用いて並置混色錯視の知覚再現の定量評価を行った. 特に, ピクセル値が異なる 2 種類の縞の配置によって同一輝度のパッチの輝度知覚が変化する Munker-White 錯視を用い, ノイズ除去とブラー除去を学習させた CAE に入力した際の出力結果を評価した. 結果, ブラー除去を学習したモデルは同時対比効果を反映する形で Munker-White 錯視の実際の知覚と輝度が逆転した結果を出力し, ノイズ除去を学習したモデルは局所的なノイズ除去学習が進行したことで錯視を知覚しなかった. また, 計算機の学習過程において錯視知覚が連続的に変化することが確認され, 錯視知覚を評価画像とすることで, 学習過程を追跡可能であることが示唆された. これらの結果から, 低レベルの視覚処理に関する学習が, 人間の視覚特性の一部に対応していることが示唆された.

今後は, 心理物理実験を行っている先行研究との比較や学習画像の変更および学習パラメータの変更に対する知覚再現の頑健性を調査することが課題として挙げられ, 低レベルの学習に関する他の処理と組み合わせることによってより汎化された錯視知覚再現の性能向上に取り組んでいきたい. 同時に, Munker-White 錯視のメカニズムとして提

案されている Oriented Differential of Gaussian (ODOG) によるモデル [16] や Contour junction による説明 [9] を取り入れた学習手法を開発し、機械学習においてこれらの視覚モデルを再現する手法についても検討したい。

#### 参考文献

- [1] 村上郁也. 視覚研究ツールとしての錯視. 光学, Vol. 39, No. 2, pp. 66–74, 2010.
- [2] A. Kitaoka and H. Ashida. Phenomenal characteristics of the peripheral drift illusion. *Vision*, Vol. 15, No. 4, pp. 261–262, 2003.
- [3] 杉原厚吉. 投影の幾何学と立体錯視. 心理学評論, Vol. 55, No. 3, pp. 296–306, 2012.
- [4] K. Shibata and S. Kurizaki. Emergence of color constancy illusion through reinforcement learning with a neural network. In *2012 IEEE International Conference on Development and Learning and Epigenetic Robotics*, pp. 1–6. IEEE, 2012.
- [5] 北崎充晃. コビキタス・イリュージョン. 日本バーチャルリアリティ学会誌= Journal of the Virtual Reality Society of Japan, Vol. 10, No. 1, pp. 13–18, 2005.
- [6] E. Watanabe, A. Kitaoka, K. Sakamoto, M. Yasugi, and K. Tanaka. Illusory motion reproduced by deep neural networks trained for prediction. *Frontiers in psychology*, Vol. 9, pp. 345–345, 2018.
- [7] A. Gomez-Villa, A. Martín, J. Vazquez-Corral, and M. Bertalmío. Convolutional neural networks can be deceived by visual illusions. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 12309–12317, 2019.
- [8] M. White. A new effect of pattern on perceived lightness. *Perception*, Vol. 8, No. 4, pp. 413–416, 1979.
- [9] B. L. Anderson. A theory of illusory lightness and transparency in monocular and binocular images: The role of contour junctions. *Perception*, Vol. 26, No. 4, pp. 419–453, 1997.
- [10] M. White. The effect of the nature of the surround on the perceived lightness of grey bars within square-wave test gratings. *Perception*, Vol. 10, No. 2, pp. 215–230, 1981.
- [11] B. Spehar, A. Gilchrist, and L. Arend. The critical role of relative luminance relations in white’s effect and grating induction. *Vision research*, Vol. 35, No. 18, pp. 2603–2614, 1995.
- [12] B. Blakeslee and M. E. McCourt. A unified theory of brightness contrast and assimilation incorporating oriented multiscale spatial filtering and contrast normalization. *Vision research*, Vol. 44, No. 21, pp. 2483–2503, 2004.
- [13] M. Cimpoi, S. Maji, I. Kokkinos, and A. Mohamed, S. and Vedaldi. Describing textures in the wild. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 3606–3613, 2014.
- [14] V. Jain and S. Seung. Natural image denoising with convolutional networks. In *Advances in neural information processing systems*, pp. 769–776, 2009.
- [15] J. Pan, S. Liu, D. Sun, J. Zhang, Y. Liu, J. Ren, Z. Li, J. Tang, H. Lu, Y. Tai, and M. Yang. Learning dual convolutional neural networks for low-level vision. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 3070–3079, 2018.
- [16] B. Blakeslee and M. E. McCourt. A multiscale spatial filtering account of the white effect, simultaneous brightness contrast and grating induction. *Vision research*, Vol. 39, No. 26, pp. 4361–4377, 1999.