

深層学習モデルを用いた MT 法によるキャラクターデザインの類似性評価

Character design similarity evaluation using MT method with deep neural network model (pre-trained model)

越水 重臣^{1*} 小黒 諒¹

Shigeomi Koshimizu^{1*} Ryo Oguro¹

¹ 東京都立産業技術大学院大学 Advanced Institute of Industrial Technology

*Corresponding author: Shigeomi Koshimizu, koshi@aiit.ac.jp

Abstract The objective of this study was to assess the degree of similarity of AI-generated character images (signal data) to existing character images (unit space data) based on the Mahalanobis distance calculated with the MT method in quality engineering. The characteristic of this study was that the feature vectors that were generated in each layer when images were input into EfficientNet-B4 used as a pre-trained model (deep neural network model) were automatically used as the feature items of the MT method for calculating the Mahalanobis distance. As a result of assessing the character images generated by AI (Midjourney) based on the keyword, “Pokémon,” using the Mahalanobis distance, we ascertained the probability with which that we could assess the similarity to existing Pokémon images.

Keywords ai-generated images; similarity; quality engineering; mt method; mahalanobis distance

1 はじめに

研究の背景

近年、生成 AI が大きな注目を集めている。今後、生成 AI はあらゆる場面で人間の仕事のやり方を変えていくであろう。例えば、デザイナーのようなクリエイターの仕事においては、まず生成 AI に基本的なデザイン案を創出させて、そこに人間であるデザイナーが自身のクリエイティビティを加えて作品を仕上げていくといった業務プロセスが想像できる。また、実際にそのようなデザインプロセスが採用され始めている[1]。

さらに、そのようなデザインプロセスにおいて、AI が複数案を生成した場面で、どのデザインが一番所望するものに近いかを数値化して評価できるとデザイナーの意思決定をより一層支援できるのではないかと考えた。すなわち、AI のした仕事を別の AI が評価するということであり、近い将来にそういったことが現実になりそうである。

このような背景から、パターン識別の手法である品質工学の MT システム[2, 3]を使った AI 生成画像の評価方法を考えたときに次の着想を得た。すなわち、AI が生成した画像が既存作品の画像に比べて類似性が高いのか／低いのかを品質工学の MT 法により数値化して評価するという方法を本報では提案する。

研究の目的および方法

本研究では、事例としてキャラクターのデザインを取り上げる。AI が生成したキャラクター画像が既存のキャラクター画像に対してどの程度の類似性を有しているかを MT 法のマハラノビス距離により数値化して評価することを目的としている。そして、具体的な題材としてゲームキャラクターの「ポケモン」を用いて研究を行う。

AI が生成したキャラクター画像は、画像生成 AI である Midjourney に「ポケモン」、「ほのおタイプ」といったキーワードをプロンプトとして与えて生成させた。プロンプトのみを使った生成方法で、既存のキャラクター画像を読み込ませるなどの操作はしていない。その結果、Midjourney が生成したキャラクター画像の 5 個を図 1 に示す。

そして、既存のポケモン画像を MT 法の単位空間とし、画像生成 AI が生成したキャラクター画像（図 1）を MT 法の信号

データとする。単位空間とは基準となるデータ群のことであり、信号データとは評価対象サンプルのことである。MT 法では、信号データと呼ばれる評価対象サンプルについて、単位空間からのマハラノビス距離を計算し、その距離の大小により、信号データが単位空間に属するか否かを判定する。具体的には、AI によって生成されたキャラクター画像が既存のキャラクター画像に似ているとなれば、MT 法で計算されるマハラノビス距離は小さくなる。すなわち、マハラノビス距離が小さいほど、その AI 生成画像は既存画像との類似性が高いと判断できるということになる。

さらに、人間が図 1 に示す AI 生成画像を見たときに感じる「ポケモンらしさ」について、アンケートによる主観評価を実施し、MT 法により計算されたマハラノビス距離とアンケート結果の相関関係を確認する。もし、両者の結果に相関があれば、MT 法で計算されたマハラノビス距離は、人間が感じる「ポケモンらしさ」を測る代用特性となりうると考えられる。



図 1 AI により生成したキャラクター画像

2 深層学習モデルを用いた MT 法

本研究で用いる事前学習済みの深層学習モデル

機械学習による画像分類では、深層学習が用いられる。深層学習では、ラベル（教師信号）を付けた画像データを集めて学習させるのだが、それが極めて手間のかかる作業となる。そのため、一般的かつ大規模な画像データセットによって既に学習

を終えた「事前学習済みモデル」を利用して別の問題に転用する「転移学習」が提案されている。事前学習済みモデルは汎用的な AI ということができる。本研究では、この事前学習済みモデルとして既存研究[4]でも使用されている EfficientNet を用いた。EfficientNet は 2019 年に Google の研究者が開発し公開した画像分類モデルで、少ないパラメータ数で高い精度を実現することが特徴である。

事前学習済みモデルを用いた特徴量の生成

EfficientNet は、EfficientNet-B0 から EfficientNet-B7 までの 8 つのモデルで構成されているが、本研究では、EfficientNet-B4 と呼ばれるアーキテクチャを採用している。その概念図を図 2 に示す。EfficientNet-B4 は 9 つの層から構成されるディープニューラルネットワークモデルである。このモデルに画像を入力すると画像データが層を通過するごとに特徴ベクトルが算出される。図 2 に示すように、特徴ベクトルの数は第 1 層で 48 個、第 9 層では 1792 個となっている。本研究では、この特徴ベクトルをそのまま MT 法の特徴項目として採用し、各層のマハラノビス距離を計算する。

これまで一般に、特徴量の設計は人間により行われてきた。キャラクターの外見的特徴量の抽出に関しても従来研究が存在し、特徴量を計測するアルゴリズムが提案されている[5]。

一方、本研究では、事前学習済みモデルが生成する特徴ベクトルをそのまま機械的に MT 法の特徴項目として採用するので、技術者が自ら特徴項目を考案する必要がないということになる。これは特徴量の考案にかかる労力がなくなるほか、別途、特徴量を計算する必要なくなるため、トータルの計算時間も短くなるというメリットにもつながる。

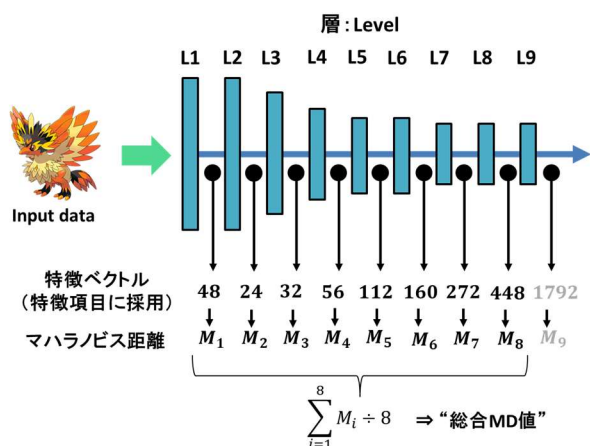


図 2 事前学習済みモデルと総合 MD 値の計算

総合 MD 値の導入

前述のようにディープニューラルネットワークの各層を通過するときに計算される特徴ベクトルを使ってマハラノビス距離を計算する。ただし、第 9 層 (L9) では特徴ベクトルの数が 1792 個もあり、後述する単位空間サンプルの数を上回るた

めマハラノビス距離が計算できない。そこで図 2 に示すように、第 1 層 (L1) から第 8 層 (L8) までで計算された 8 個のマハラノビス距離の平均値を計算して、これを「総合 MD 値」と呼び、本研究における評価指標とした。

3 ジャックナイフ法による単位空間の調査

一つ抜き交差検証による解析方法

ここでは、ジャックナイフ法 (一つ抜き交差検証) を用いて、単位空間サンプルの調査を行う。一つ抜き交差検証とは、1 個のデータのみをテスト用に利用し、残りを全て学習データとして利用する交差検証のことを言う。本研究の MT 法における一つ抜き交差検証の方法を図 3 に示す。単位空間は 1190 サンプルのポケモン画像である。画像は「ポケモンずかん」のサイト [6] のものを採用した。まず、単位空間から 1 つのサンプル画像を抜き取る。この時、単位空間に属するサンプル数は 1189 になる。これを新たに単位空間として、抜き取った 1 サンプルについて総合 MD 値を計算する。続いて、抜き取ったサンプルを単位空間に戻し、また別のサンプルを 1 つ抜き取り、総合 MD 値を計算する。以下、単位空間の 1190 サンプルすべてについて同様の計算を繰り返し、1190 サンプルの総合 MD 値を得る。この調査により 1190 サンプルの総合 MD 値が得られることになり、最もポケモンらしいキャラクターデザインからそうでないものまでを総合 MD 値の順番に並べることができる。



図 3 一つ抜き交差検証による単位空間の調査方法 (画像はポケモンずかん[6]より引用)

一つ抜き交差検証による解析結果

ジャックナイフ法による解析結果を以下に示す。まず、総合 MD 値が小さくなったポケモン画像群 (上位 5 個) を図 4 (a) に示す。総合 MD 値が小さいということは「ポケモンらしい」キャラクターデザインであることを意味している。同図の画像下の数値は総合 MD 値を示しているが、左端の画像が最小値の 0.42 となっている。5 個の画像がいずれも丸みを帯びたかわいい感じのキャラクターになっていることが視認できる。逆に総合 MD 値が大きくなったポケモン画像群 (上位 5 個) を図 4 (b) に示す。総合 MD 値が大きいは「ポケモンらしくない」キャラクターデザインを意味していると言える。同図では右端画像の総合 MD 値が最大値である 10.69 となっている。図 4 (a) の画像と比べると、例えばギザギザとした鋭角なデザイン要素が含まれるなど、明らかにキャラクターデザインは変化してお

り、ポケモンキャラクターとしてはレアな印象を受ける。

このようにジャックナイフ法により単位空間に属する全 1190 サンプルの画像について、総合 MD 値による順位付けをすることができた。さらに、その結果をヒストグラムで表したものが図 5 である。ヒストグラムに累積比率を示す折れ線グラフを重ねているが、そのグラフによれば、総合 MD 値が 3 を超えるようなサンプルの割合は全体の 1.2% と小さい割合であることがわかった。すなわち、総合 MD 値が 3 を超えるようなポケモン画像は、かなりレアなキャラクターデザインと判断でき、ポケモンらしさを判別するための 1 つの基準となるのではないかと考えられる。



(a) 総合 MD 値が小さいポケモンキャラクター



(b) 総合 MD 値が大きいポケモンキャラクター

図 4 ジャックナイフ法による解析結果（画像はポケモンずかん [6] より引用）

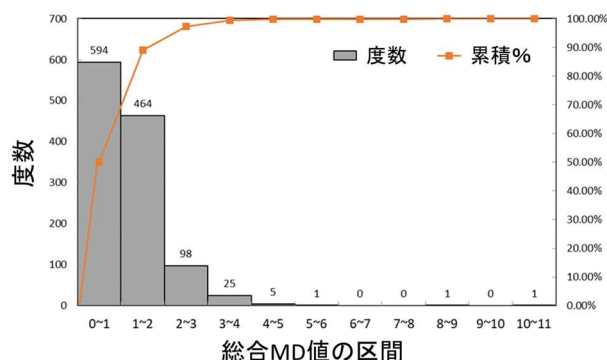


図 5 単位空間サンプルのヒストグラムと累積比率

4 総合 MD 値による AI 生成画像の類似性評価

解析方法

ここでは、総合 MD 値を使って図 1 の AI 生成画像を評価する。その解析方法の概略を以下にまとめる。

・単位空間：既存のポケモン画像 1187 サンプル（ピカチュウ、ヒトカゲ、リザードンは含まれない）

・信号データ：AI が生成したキャラクター画像 5 サンプル（図 1）とピカチュウ、ヒトカゲ、リザードン

・評価方法：信号データの総合 MD 値を計算し、単位空間からの距離が小さい画像ほどポケモンらしい、すなわち、既存画像との類似性が高いと判断する

ちなみに、信号データには、ポケモンキャラクターの代表として、ピカチュウ、ヒトカゲ、リザードンをベンチマーク的な意味合いで加えている。今回、Midjourney を使って生成したキャラクター画像は、その生成時のプロンプトに Pokémon のほか Flame type（ほのおタイプ）の文言を記述しており、ほのおタイプのポケモンを生成したことになっている。そのため、ほのおタイプのポケモンの代表としてヒトカゲとリザードンを加えている。

解析結果

事前学習済みモデルの EfficientNet-B4 に信号データを入力する。EfficientNet-B4 おける 8 つの層を通過するとき計算される特徴量ベクトルを MT 法の特徴項目に採用し各層のマハラノビス距離を計算した。その結果を図 6 に示す。

さらに、それら 8 個のマハラノビス距離の平均値である総合 MD 値を用いて、信号データを比較した結果を図 7 に示す。図 7 では左から総合 MD 値が小さくなる画像の順番に並べている。また図 6 と図 7 には、参考として単位空間サンプルの平均値（UnitSpace_Ave.）も示している。

図 7 では、その総合 MD 値が小さいほど単位空間に近いということになる。すなわち、既存ポケモンキャラクターのデザインに関する特徴をよく捉えており、よりポケモンらしい画像であると言える。実際に、ピカチュウ、ヒトカゲ、リザードンはその総合 MD 値が小さくなっている。順番としては、それらの後に 5 つの AI 生成画像が続くわけであるが、その中でも総合 MD 値が小さい方がよりポケモンらしい画像になっていると判断できる。例えば、前章で述べた総合 MD 値の閾値 3 を採用して、図 7 の AI 生成画像を見てみると、AI_3 と AI_0 の 2 つの画像は総合 MD 値が閾値 3 を下回っており、ポケモンらしいデザインであると判別できる。逆に、AI_1、AI_4、AI_2 の 3 つの画像については、閾値 3 を超えており、ポケモンらしくないデザインであると判別できる。

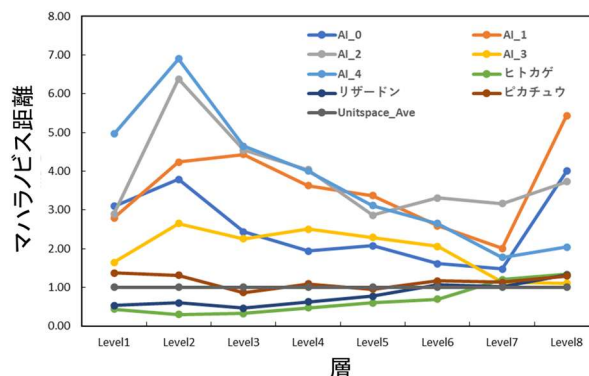


図 6 DNN の各層におけるマハラノビス距離

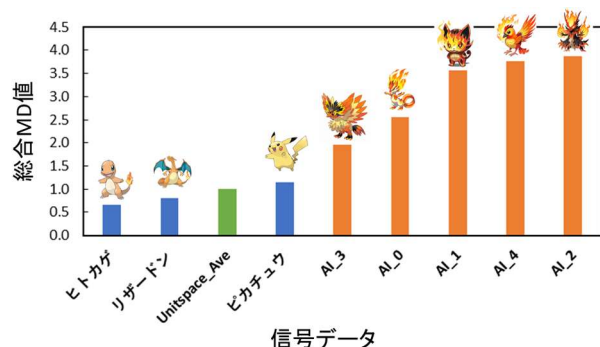


図 7 総合 MD 値による信号データの評価（一部の画像はポケモンずかん[6]より引用）

6 おわりに

本研究の目的は、AI が生成したキャラクター画像（信号データ）が既存のキャラクター画像（単位空間データ）に対してどの程度の類似性を有しているかを MT 法のマハラノビス距離により数値化して評価することである。そして本研究の特色は、ディープニューラルネットワークの事前学習済みモデルである EfficientNet-B4 に画像を入力したとき、各層で生成される特徴量ベクトルをそのまま MT 法の特徴項目に採用し、マハラノビス距離を計算することである。

本報で述べた総合 MD 値により、ポケモンというキーワードをもとに AI（Midjourney）が生成したキャラクター画像を評価したところ、既存のポケモン画像との類似性を数値化できている可能性を得た。AI 生成画像を人に見てもらい「ポケモンらしさ」を評価したアンケート結果と MT 法の総合 MD 値との間に相関が見られたことから、総合 MD 値による評価は人間が感じるポケモンらしさをうまく数値化できているのではないかと考えられる。

7 謝辞

東京都立産業技術大学院大学の修了生である大坂冬子氏には本研究に関して貴重な議論や助言をいただいた。同じく、同大学の修了生である渡邊泰成氏には研究遂行にあたり多大な協力をいただいた。ここに記して深く感謝する。

参考文献

1. 日経デザイン編. ヒット商品の裏にデザインあり！ (日経 BP ムック). 日経 BP; 2024.
2. Shoichi T, et al. QUALITY RECOGNITION AND PREDICTION Smarter Pattern Technology with the Mahalanobis-Taguchi System, MOMENTUM PRESS; 2012.
3. 田口玄一監修. 品質工学便覧 第 5 章 MT システムのパターンの診断と SN 比-MT 法-, p128-134, 日刊工業新聞社; 2007.
4. Rippel O, et al., Modeling the Distribution of Normal Data in Pre-Trained Deep Features for Anomaly Detection, (2020), <https://arxiv.org/abs/2005.14140>
5. 高松耕太, 島津恵子. キャラクターの外見的特徴量の計測実験. 情報処理学会研究報告. 2010 年度 (5), 1-4, 2011-02
6. ポケモンずかん, ©2025 Pokémon. ©1995-2025 Nintendo/Creatures Inc./GAME FREAK inc., <https://zukan.pokemon.co.jp/>

5 アンケートによるポケモンらしさの主観評価

アンケートの方法

図 1 に示した AI 生成画像を被験者に見てもらい、「ポケモンらしさ」があると思うかを問うアンケートを実施した。5 つの AI 生成画像について、以下に示す 4 段階で回答してもらい対応する評点を得た。

- ・そう思う ⇒ 評点：4
- ・どちらかと言えばそう思う ⇒ 評点：3
- ・どちらかと言えばそう思わない ⇒ 評点：2
- ・そう思わない ⇒ 評点：1

アンケート回答数は n=32 であり、AI 生成画像ごとに評点の平均値を求めた。

アンケートの結果

アンケートの結果を表 1 に示す。同表上段は評点平均値の高い順に左から並べている。また参考までに、同表下段には 4 章で述べた AI 生成画像の総合 MD 値による順位も付記している。アンケート評点と総合 MD 値の両方を比較すると、3 位以下のところでは順位が入れ替わっているところがあるものの、1 位と 2 位については同じ評価結果をなっていることがわかる。したがって、総合 MD 値による評価は、人間が感じるポケモンらしさを数値化することにおおよそ成功しているのではないかと考えられる。

表 1 アンケート結果

	1位	2位	3位	4位	5位
アンケート 評点平均値 (n=32)	 3.8	 3.6	 2.3	 2.1	 1.5
総合MD値	 2.0	 2.6	 3.6	 3.8	 3.9