

# 深層学習を用いたスポーツ画像学習モデルによる スポーツピクトグラムの認識手法の提案

佐野 景飛<sup>†</sup> 清 雄一<sup>†</sup> 田原 康之<sup>†</sup> 大須賀 昭彦<sup>†</sup>

<sup>†</sup> 電気通信大学 〒182-8585 東京都調布市調布ヶ丘 1-5-1

E-mail: <sup>†</sup> sano.keito@ohsuga.lab.uec.ac.jp, {seiuny, tahara, ohsuga}@uec.ac.jp

**あらまし** ピクトグラムとは、不特定多数の人々が利用する公共交通機関や公共施設、観光施設において、文字・言語によらず対象物、概念または状態に関する情報を提供する図形のことであり、近年では東京オリンピックが開催されるなど、多くの人々がスポーツピクトグラムに触れる機会が増えてきている。しかし、ピクトグラム単体で見た際に何の競技を表しているのかわかりづらいものがいくつか存在するため、全てのピクトグラムに対しその概念を完全に理解することは困難である。そこで、本論文では実際のスポーツ画像を学習させたモデルをピクトグラムでファインチューニングするという手法を提案した。100種類のスポーツに対し合計14600枚の実際のスポーツ画像で構成されたデータセットで学習させ、ピクトグラムでファインチューニングすることで、高い精度を実現することを確認した。

**キーワード** 画像認識, 深層学習, スポーツ, ピクトグラム

## 1. はじめに

ピクトグラム（案内用図記号）とは、不特定多数の人々が利用する公共交通機関や公共施設、観光施設等において、文字・言語によらず対象物、概念または状態に関する情報を提供する図形[1]のことである。近年では東京オリンピックが2021年に開催されるなど、多くの人々がピクトグラムに触れる機会が増えてきている。

オリンピックでは、開催されるたびに行われる各競技のピクトグラムのデザインが一新されることを例に昨今では様々なところで日々新しいピクトグラムが作成されている。しかし、中にはピクトグラム単体で見た際に何の競技を表しているのかわかりづらいものや、デザインが似ているためわかりづらいものがいくつか存在する。図1の出典は2020東京オリンピックで使用されたピクトグラムである。[2]



図1 分かりづらいピクトグラムの例

また、吉田らの研究[3]や神田らの研究[4]では、深層学習を用いたピクトグラム画像に対する画像分類問題を扱っている。吉田らの研究では、案内所、情報コー

ナー、お手洗い、男子、女子、障がい者用、喫煙、エレベーター、エスカレーター、階段の10種類のピクトグラム画像の分類を行っており。神田らの研究では、JIS Z8210で規定されている180種類のピクトグラムに対して高い精度で分類を行っている。これら先行研究の学習データの作成方法については、学習に使うピクトグラム（RGB3チャンネル、画像サイズは縦横共に70ピクセル）に対し、それぞれ回転、透視変換、平行移動、左右反転という画像処理を行い、学習データとテストデータを作成している。更に、神田らの研究では、事前に別のタスクに向けて学習した畳み込みニューラルネットワークのパラメータを初期値としたファインチューニングを行っている。具体的には、ImageNet[5]で学習済みのVGG16[6]モデルの情報を利用することで、畳み込みニューラルネットワークの特定の層までの重みのデータを固定し、それ以降の層においては重みを再学習させるという手法をとっている。しかし、認識するピクトグラムのデザインと、対応するImageNetの画像が極端に異なる場合が多く、ファインチューニングを最大限に活用できていないのではないかとこの疑問がある。よって、本研究の目的は、ディープラーニングと呼ばれる機械学習の手法を用い、実際のスポーツシーン画像を学習させたモデルを適切にピクトグラムでファインチューニングし、スポーツピクトグラムの分類に対して精度を上げることで、スポーツピクトグラムのデザインのみが与えられた際、そのピクトグラムが何を表しているのかを正しく分類することによって利用者貢献することである。

## 2. 提案手法

本研究では、研究テーマにもある通り深層学習を用いてディープニューラルネットワークを組むことで画像分類問題を扱う。そこで、「スポーツを行う際の姿勢」がピクトグラム分類の特徴になると考え、スポーツ画像を学習させたモデルをピクトグラムでファインチューニングするという手法をとる。入力に使用するピクトグラム画像はピクトグラムを中央に示した矩形のものを想定しており、出力は入力画像と認識結果を同時に出力するものとする。以下に入出力の想定例を示す。

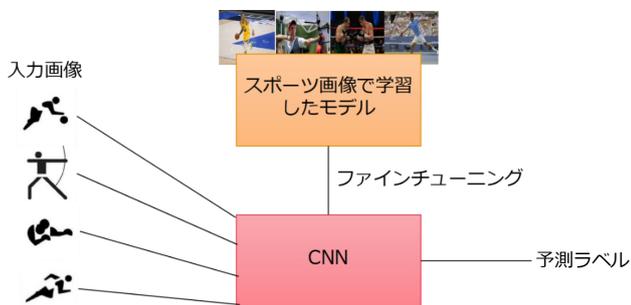


図2 システム構成図

ピクトグラムの解析にはディープニューラルネットワークを用い、ディープニューラルネットワークによる解析は畳み込みニューラルネットワーク(CNN)を用い、学習には教師ありの機械学習を用いる。使用するデータセットを構成する各ピクトグラムは、正解の競技(ラベル)の名前のついたフォルダにそれぞれ保存することで対応させている。

## 3. 分類器

### 3.1 分類器の実装

本研究で使用するCNNの実装はPyTorchを利用して行った。CNNの構造は、6層の畳み込み層全てでmaxpooling[7]を行い、入力はRGB3チャンネル、幅300ピクセル、高さ300ピクセル、バッチサイズを64とし、損失関数には交差エントロピー誤差、活性化関数にはReLUを用いた。最適化アルゴリズムには、学習率を0.01に設定したAdamW[8]を用い、エポック数を100とした。

### 3.2 スポーツ画像学習モデル

モデルを訓練するために用いたデータセットは、kaggleが提供する100種類のスポーツシーン画像を

14600枚収集した「100 Sports Image Classification」データセット[9]を用いた。入力画像の前処理としてそれぞれの画像を幅300ピクセル、高さ300ピクセルにリサイズし、各色チャンネルの平均値を0.5、標準偏差を0.5として色情報の標準化を行い、Tensor型に変換する処理を加える。

### 3.3 ファインチューニング

訓練データとテストデータは、過去に行われた夏季・冬季オリンピックで用いられた各競技のピクトグラムを、各年のデザイナー一枚ずつを収集し構成したものを使用した。

### 3.4 出力

出力は、CNNが受け取ったピクトグラムに対して予測したラベルを出力するものとする。更に、Pythonのオープンソースであり、機械学習ライブラリであるscikit-learn(バージョン0.24.2)を用いて、以下の評価指標となる数値をそれぞれ出力する。CNNによって予測されたラベルと正しいラベルの情報をリストとして入力し、「予測したラベルと正解のラベルがどれほど一致していたかを示す数値」であるaccuracy、「あるラベルを正として、正と予測した結果の正解ラベルがどれほど一致していたかを示す数値」であるprecision、「あるラベルを正として、実際に正であるものに対してどれほど正しい分類が出来たかを表す数値」であるrecall、「precisionとrecallの両方の性質を考慮し、それらの調和平均をとった数値」であるf1-scoreの4つの評価指標となる値を出力する。

## 4. 実験・評価

### 4.1 実験

スポーツ画像学習モデルをピクトグラムでファインチューニングしたモデルである提案手法を使用して実際にピクトグラム分類問題を扱う。42種類のスポーツピクトグラム計442枚を訓練データ、テストデータに設定し、ピクトグラムのみで学習したモデルと、提案手法を用意し、これら2つのモデルでピクトグラム認識の精度にどれほどの違いがみられるのか実験を行った。

### 4.2 評価

実験の結果を以下の表1に示す。ここで、「precision」、「recall」、「f1-score」については3.4項で述べたものを用い、「support」は442枚のピクトグラムの内、該当するスポーツのピクトグラムが何枚あるかの数値を表している。

表 1: 各モデルの分類精度

	ピクトグラムのみ	提案手法
accuracy	0.505	0.739
f1-score の平均	0.496	0.721

表中の「accuracy」は、予測されたラベルが正解のラベルとどれほど一致していたかを表す数値であり、「f1-score の平均」は、あるラベルを正として、正と予測した結果が一致していたかを表す指標と、実際に正であるものに対してどれほど正しい分類が出来たかを表す指標の調和平均をとった数値である。表 1 から、認識手法の指標が高いことが分かる。

図 3 各モデルの loss 比較

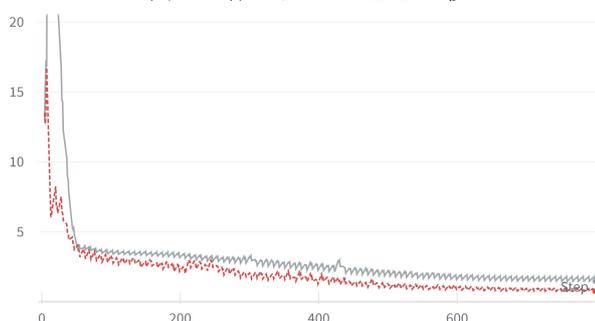


図 3 は、ピクトグラムのみで学習したモデルの loss (灰色実線) と提案手法の loss (赤点線) の減少の様子をグラフで可視化したものである。これらのグラフより、ピクトグラムのみで学習したモデルよりも提案手法のほうが loss が減少していることが分かった。

## 5. 考察

1 種類のスポーツに対するピクトグラムのデザイン数は、その競技がオリンピックで行われた回数分しか作成されないため、競技によって数に偏りが生まれる。競技ごとのピクトグラムデザインの種類が多い競技程、一つの競技における認識の回数が多くなり、正確な精度が求まる可能性が高くなると考えられる。本論文では、10 種類以上のデザインで学習することができる競技をピクトグラムデザインの種類が多い競技とし、考察を行っていく。まず、表 1 において認識精度が向上しているが、これは 10 種類未満のデザインで学習する競技における精度の改善と比べて 10 種類以上のデザインで学習する競技において精度が改善した競技が多かったためといえる。よって提案手法での認識結果は、デザインの種類が多いほど精度が改善しており、偶然認識精度が高くなったわけではないことが分かった。

## 6. おわりに

本論文のまとめとして、ディープニューラルネットワークを用いて、ピクトグラムの画像分類問題を扱ったが、ピクトグラムのみで学習したモデルよりも、本論文の提案したスポーツ画像を学習させたモデルをピクトグラムでファンチャーニングしたモデルの分類精度が高いという結果が得られた。今後の課題として、先行研究のようにピクトグラムのデータセットに対して画像処理を加えたり、デザインの種類を増やすことによってデータセットの枚数を増やすことが挙げられる。更に、この画像分類システムを実際に運用するにあたって、オリンピックで使用されたピクトグラムのみならず、データセットに存在しないピクトグラムに関しても扱うことが出来ればなお良いと考える。

## 謝 辞

本研究は JSPS 科研費 JP21H03496, JP22K12157 の助成を受けたものです。

## 参 考 文 献

- [1] 国土交通省案内用図記号 (JIS Z8210) [https://www.mlit.go.jp/sogoseisaku/barrierfree/sosei\\_barrierfree\\_tk\\_000145.html](https://www.mlit.go.jp/sogoseisaku/barrierfree/sosei_barrierfree_tk_000145.html)
- [2] Tokyo 2020 unveils Games pictograms <https://olympics.com/en/news/tokyo-2020-unveils-games-pictograms>
- [3] 吉田雄大, 伊藤一成: 深層学習を用いたピクトグラム画像への情報補完手法の提案, マルチメディア, 分散, 協調とモバイル DICO2017 シンポジウム, pp. 1075-1080(2017)
- [4] 神田剛志, 伊藤一成: 深層学習を用いたピクトグラム画像における多クラス分類問題とその応用, エンタテインメントコンピューティングシンポジウム 2019 論文集, pp.73-79(2019)
- [5] Jia Deng, Wei Dong, Richard Socher, Li-Jia Li, Kai Li and Li Fei-Fei: ImageNet: A Large-Scale Hierarchical Image Database. In 2009 IEEE Conference on Computer Vision and Pattern Recognition
- [6] Karen Simonyan and Andrew Zisserman: Very deep convolutional networks for large-scale image recognition. In ICLR2015
- [7] 宮本崇: 畳み込みニューラルネットワークにおける順伝播・誤差逆伝播計算(2017)
- [8] Ilya Loshchilov & Frank Hutter: Decoupled Weight Decay Regularization, In ICLR2019
- [9] 100 Sports Image Classification | Kaggle <https://www.kaggle.com/datasets/gpiosenka/sports-classification>