

# 深層学習に基づく画像認識技術を用いた 部屋の片付き具合の自動推定

藤平 翔† 岩野 公司†

東京都市大学†

## 1. はじめに

部屋を片付けたいと思いながらも、散らかった状態のまま過ごしてしまう人がある。片付けを行うきっかけの一つとして、「他者からの指摘や評価を意識する」ことが挙げられる。そこで、情報システムによって部屋の様子をモニタリングし、その部屋の状態の評価を本人に提示することができれば、それが動機となって片付けが進むことが期待される。

このような観点から、我々は部屋の写真の画像からその部屋の「片付き具合」を評価する手法の開発を進めている[1]。先行研究[1]では、部屋画像から各種の画像特徴量（検出物体数や色のエントロピーなど）を抽出し、重回帰分析で求めた回帰式を利用して片付き具合を推定する手法を提案している。本研究では、推定性能の向上を目指し、深層学習に基づく畳み込みニューラルネットワーク（CNN: Convolutional Neural Network）を利用した部屋の片付き具合の推定手法を提案し、その評価を行う。

## 2. 学習に使用する画像データ

実験に使用する画像データとして、45枚の部屋画像を用意する。対象とした部屋は、寝具、机、テーブルなどがあるワンルームに限定した。これらの画像は、縦の画素数が400に統一されるようにサイズの調整を行っている。

各画像に対し、大学生20名により、部屋の片付き具合を5段階（1:絶対に片付けを行うべき～5:まったく片付ける必要がない）で評価してもらった。この人間の平均評価値を自動推定することが目的となる。なお、45枚の画像のうち、3.0以上となる「高評価」の画像は25枚、3.0未満になる「低評価」の画像は20枚であった。実験に使用する部屋画像の例を図1に示す。

## 3. 深層学習による片付き具合推定手法の提案

### 3.1 ネットワークの学習方法

提案する手法では、以下のステップで片付き

Automatic estimation of room tidiness using deep-learning-based image recognition technique  
Sho Fujihira†, Koji Iwano†, †Tokyo City University



図1 使用する部屋画像の例（平均評価値：2.9）

具合推定用のCNNの学習を行う。

1. 初期ネットワークとして、画像認識用の大規模データベース ImageNet[2]で学習されたCNNを利用する。
2. 初期ネットワークの全結合層と出力層を取り除いた上で、1層の全結合層（活性化関数: ReLU）と2ノードの出力層（活性化関数: ソフトマックス関数）を追加し、「高評価」と「低評価」の2クラス分類を行うネットワークに変更する。その上で、部屋画像の学習データを用いて全結合層と出力層のパラメータを転移学習により最適化する。このときの損失関数には交差エントロピー誤差を用いる。
3. 2で得られたネットワークに1ノードの出力層（活性化関数: 線形関数）をさらに追加し、人間の平均評価値の予測するネットワークに変換する。それを初期ネットワークとして、学習データを用いて全結合層以降を最適化する。このときの損失関数には平均二乗誤差を使用する。

なお、CNNにはVGG16[3]とResNet50[4]を使用し、性能の比較を行う。

### 3.2 学習データの選択と拡張

3.1節のステップ2の学習を行う際に、学習データの選択と拡張の適用を検討する。

データ選択の方法としては、高評価画像を平均評価値4.0以上、低評価画像を平均評価値2.0未満となるものに限定する。45枚中、前者は13枚、後者は10枚である。このようにすることで、中間的な画像が取り除かれて判定の難易度が下がることから、このデータの利用によりカリキュラムラーニングが導入できる。

表 1 提案手法（ステップ 2 まで）における 2 クラス分類性能の評価結果

手法	データ選択・拡張		正解率 (%)
	1 段階目	2 段階目	
重回帰分析による推定[1]			55.6
VGG16	TA	—	75.6
	WA, TA	—	68.9
	DS, WA, TA	—	77.8
	WA, TA	TA	73.3
	DS, WA, TA	TA	<b>84.4</b>
ResNet50	TA	—	71.1
	WA, TA	—	71.1
	DS, WA, TA	—	75.6
	WA, TA	TA	75.6
	DS, WA, TA	TA	<b>86.7</b>

学習データ不足を補うためのデータ拡張として、2 種類の方法の適用を考える。1 つ目は、ウェブ検索[5]で収集した類似画像の追加で、この操作ではデータが約 24 倍に拡張される。2 つ目は、画像変換（左右反転と輝度値の変更）に基づくデータ拡張で、データが 6 倍に拡張される。なお、学習ステップ 3 においては、画像変換によるデータ拡張のみを適用する。

#### 4. 推定性能の評価結果

##### 4.1 2 クラス分類の性能検証

ステップ 2 で構築される 2 クラス分類の CNN について性能評価を行う。評価は 9 分割の交差検証で実施した。表 1 に、データ選択 (DS)、ウェブ検索によるデータ拡張 (WA)、画像変換によるデータ拡張 (TA) の適用の組み合わせを変えて評価を行った場合の分類の正解率を示す。50 エポックで学習を行った場合と、それに続けて (データ選択・拡張の方法を変えて) さらに 50 エポック追加して (2 段階で) 学習した場合があり、それぞれの性能を示している。最適化アルゴリズムには Adam[6]を用い、バッチサイズは 10 に統一している。参考のため、先行研究[1]の重回帰式による推定値で 2 クラス分類を行った場合の正解率も示す。

この結果をみると、(1) DS の導入により分類性能が向上すること、(2) WA で増えたデータは、被験者が実際に評価をしておらず、ラベルの精度が悪いため、WA を利用しない追加学習が必要であることが確認できる。また、先行研究の重回帰式による推定よりも CNN を用いた提案手法の方が高い性能になっていることもわかる。

##### 4.2 部屋の片付き具合予測の性能検証

2 クラス分類において、VGG16, ResNet50 のそれぞれで最高性能だったネットワークを使用し、

表 2 提案手法（ステップ 3 まで）における部屋の片付き具合の推定誤差の評価結果

手法	平均二乗誤差
重回帰分析による推定[1]	1.53
VGG16 による推定	1.06
ResNet50 による推定	1.01
人間による評価	0.94

学習ステップ 3 を実行して、部屋の片付き具合を予測するネットワークを構築し、その推定誤差 (平均二乗誤差) を求めた。この実験も 9 分割の交差検証により実施している。学習時の際の最適化アルゴリズムには Adam[6]を用い、バッチサイズは 10、エポック数は 50 とした。

表 2 に結果を示す。参考のため、先行研究[1]の重回帰式による推定の誤差と、人間 (20 名の被験者) が評価を行った際の平均評価値との誤差も示す。これを見ると、ResNet50 の方が VGG16 よりも性能が高く、人間の評価誤差に迫っていることがわかる。また、提案手法の方が先行研究[1]の重回帰式による推定より性能が高いことも確認できる。

#### 5. まとめ

本研究では、部屋の片付き具合 (人間が評価したときの平均評価値) を、CNN により推定する手法の提案を行った。具体的には、事前学習された CNN を 2 クラス分類器として転移学習した上で、線形予測を行うネットワークに変形し、さらに転移学習を行う方法である。実験の結果、ResNet50 を用いた場合に良好な性能となることがわかった。

今後の課題としては、さらなる推定性能の改善を図るため、データの拡充や Vision Transformer などの利用を検討する必要がある。

#### 参考文献

- [1] 藤平, 岩野, “画像処理技術を用いた部屋の片付き具合の自動推定,” 電子情報通信学会総合大会, ISS-SP-014, p. 148, 2022.
- [2] <https://www.image-net.org>
- [3] K. Simonyan and A. Zisserman, “Very Deep Convolutional Networks for Large-Scale Image Recognition,” Proc. International Conference on Learning Representations, 2015.
- [4] K. He, et al., “Deep Residual Learning for Image Recognition,” Proc. IEEE Conference on Computer Vision and Pattern Recognition, pp. 770-778, 2016.
- [5] Google 画像検索, <https://www.google.co.jp/imghp>
- [6] D. P. Kingma, J. Ba, “Adam: A Method for Stochastic Optimization,” Proc. ICLR, 2015.