

写真画像に基づく自動用紙推定の提案

Automatic Photographic Paper Estimation based on Photographic Images

平井喜一^{1*} 上野未貴¹
Kiichi Hirai¹ Miki Ueno¹

¹ 大阪工業大学

¹ Osaka Institute of Technology

Abstract: 写真愛好家らが写真を印刷する際、印刷物の印象は使用する用紙で変わり、創作物の表現に大きな影響を与える。そのため、写真に調和する用紙を選ぶ作業は難しく、用紙選択に迷うことがある。現状はコンテスト応募時などに創作者が用紙を決め印刷しており、自動化の事例は未だない。本研究では、写真を元に調和する写真用紙を機械学習させ、印刷時に推定・提示し、創作支援をするシステムを提案する。

1 はじめに

写真愛好家らが写真を作品制作やフォトコンテストへの応募のために印刷する際、印刷物の印象は使用する用紙の見え方によって変わり、創作物の表現に大きな影響を与える [1]。また、写真用紙は光沢紙やマット紙など様々な種類が販売されており、写真に調和する用紙を選ぶ作業は非常に難しい。よって、用紙選択に迷うことがある。現状は創作者の経験や推量で行なっているため、初心者などの経験がない人は写真に対してどのような用紙が適しているのかわからず挫折してしまう人が多くいる。そこで本研究では、写真を元に調和する写真用紙を正解ラベルとした写真画像を機械学習し、用紙推定モデルを作成する。また、作成したモデルを用いて創作者が印刷をする際に最適な用紙を推定し提示させるシステムを開発する。さらに写真のどの部分を元に用紙を推薦したかを可視化させることで印刷技術向上につなげる。

2 要素技術

2.1 畳み込みニューラルネットワーク

本研究では、写真に調和する写真用紙を学習させるために、画像認識の分野で優れた性能を発揮している畳み込みニューラルネットワーク (Convolutional Neural Network: CNN) を利用し、大規模画像データセットである ImageNet [2] の一般物体認識データセットで事前

学習済の VGG-16 モデル [3] を使用する。VGG-16 モデルは 2014 年の ILSVRC で提案され、13 層の畳み込み層、その途中に挟まれている 5 層のプーリング層、3 層の全結合層および出力層から成り立っているモデルである。

2.2 ファインチューニング

CNN を利用し高い精度でモデルを作成するには、膨大な学習・評価データが必要でありデータセット構築が困難である。そこで、少数画像からモデルを構築する手法として転移学習やファインチューニングなどの手法がある。転移学習では事前に学習を行なったモデルの出力層以外を重みデータは変更せずに特徴抽出器として利用し、出力層に近い数層のみを学習させることで新しいモデルを構築する。対してファインチューニングでは、事前に学習を行なった出力層以外のモデルの重みデータを一部再学習して特徴量抽出器として利用し、新しいモデルを構築する [4]。

2.3 Grad-Cam

Gradient-weighted Class Activation Mapping (Grad-Cam) [5] とは、CNN の判断根拠の可視化技術である。Grad-Cam は CNN の最終層の畳み込み層により抽出された特徴量に着目して、CNN が画像のどの部分を見ているのかを可視化する技術である。

*連絡先：大阪工業大学 工学部 電子情報通信工学科
〒535-8585 大阪府大阪市旭区大宮 5-16-1
E-mail: e1617085[at]st.oit.ac.jp

表 1: 種類別枚数

種類	用紙	枚数	総枚数
光沢紙	フォトペーパー	193	361
	パールラベルペーパー	58	
	ホワイティフィルムペーパー	110	
半光沢紙	セミグロスペーパー	362	599
	ブルーラベルペーパー	59	
	レッドラベルペーパー	18	
	シルバーラベルペーパー	144	
	フォトキャンバスペーパー	16	
マット紙	デザインペーパー	22	160
	コットンペーパー	12	
	ソフトグロスペーパー	106	
	スムーズフォトペーパー	20	

3 提案手法

3.1 概要

本研究では、用紙推定のために CNN を用いて写真に調和する写真用紙を正解ラベルとした写真画像を機械学習させ用紙推定モデルを作成し、モデルを用いて任意の画像を元に推論結果と分類寄与箇所を出力する手法を提案する。

3.2 データセット

写真画像と調和する用紙を正解ラベルとしたデータセットは存在せず、本研究では用紙推定モデルを構築するためのデータセットとして、ダイヤモンド株式会社主催している「ピクトリコフォトコンテスト」の入賞作品¹を用いた。写真に調和する用紙は各個人の感性によって変わり正解を決めるのが困難である。しかし、プロ写真家が選出した入賞作品を用いることで写真画像と用紙の組合せは調和しており正解であると考えられる。本データセットは 2011~2020 年までの 9 年間で 12 種類の写真用紙に印刷されている 1120 枚の写真を集め、さらに用紙情報に基づいて光沢紙・半光沢紙・マット紙に分類した。表 1 に種類別枚数を示す。

3.3 モデル作成

CNN は通常、画像分類に使用される。本研究では、写真に調和する用紙を正解ラベルとした写真画像を入力、調和する用紙を出力として CNN を用いて分類モデルを作成する。提案手法では事前学習済み VGG-16 モデルの畳み込みの 12 層目以降について再度学習を行うファインチューニングを適用し、出力層を 3 クラスにした。

¹<https://www.pictorico.co.jp/system/contents/2105/>

表 2: 学習器のパラメータ

パラメータ	値
Batch Size	32
Epochs	100
最適化関数	SGD
学習率	0.005
損失関数	交差エントロピー

表 3: CNN ネットワークパラメータ

層	出力次元	フィルタサイズ	活性化関数	Dropout
Input	224 × 224 × 3	-	-	-
Conv2D	224 × 224 × 64	3 × 3	ReLU	-
Conv2D	224 × 224 × 64	3 × 3	ReLU	-
MaxPooling2D	112 × 112 × 64	2 × 2	-	-
Conv2D	112 × 112 × 128	3 × 3	ReLU	-
Conv2D	112 × 112 × 128	3 × 3	ReLU	-
MaxPooling2D	56 × 56 × 256	2 × 2	-	-
Conv2D	56 × 56 × 256	3 × 3	ReLU	-
Conv2D	56 × 56 × 256	3 × 3	ReLU	-
Conv2D	56 × 56 × 256	3 × 3	ReLU	-
MaxPooling2D	28 × 28 × 256	2 × 2	-	-
Conv2D	28 × 28 × 512	3 × 3	ReLU	-
Conv2D	28 × 28 × 512	3 × 3	ReLU	-
Conv2D	28 × 28 × 512	3 × 3	ReLU	-
MaxPooling2D	14 × 14 × 512	2 × 2	-	-
Conv2D	14 × 14 × 512	3 × 3	ReLU	-
Conv2D	14 × 14 × 512	3 × 3	ReLU	-
Conv2D	14 × 14 × 512	3 × 3	ReLU	-
MaxPooling2D	7 × 7 × 512	2 × 2	-	-
Flatten	-	-	-	-
Dense	256	-	ReLU	-
Dropout	-	-	-	0.5
Dense	3	-	Softmax	-

4 実験

4.1 実験条件

学習は、GPU(NVIDIA Quadro RTX6000)を搭載した Ubuntu18.04.5LTS 上の Python3.7.6 でバックエンドに TensorFlow-GPU2.1.0 を用いた Keras2.4.2 でした。

4.2 分類

表 1 で分類した 3 種類(クラス)の各枚数を、総枚数が一番少ないマット紙に合わせるためランダムに減らし、すべてのクラスを 160 枚で揃えた。次に、学習・評価・テストデータ=8:1:1 に分け、学習データに対して左右・上下反転のデータ拡張をし、事前学習済みの VGG16 モデルを用いてファインチューニングをした。表 2,3 に実験で使用したパラメータを示す。最後にテストデータを用いてモデルの汎化性能を確認し Grad-Cam で特徴量可視化をした。

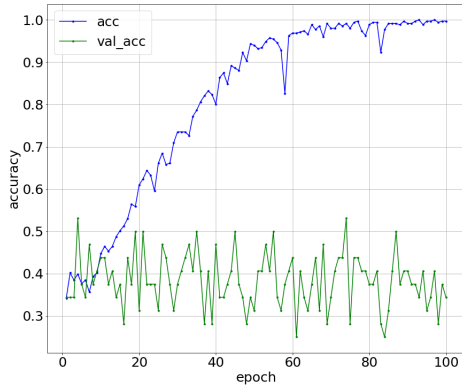


図 1: 認識精度の変化

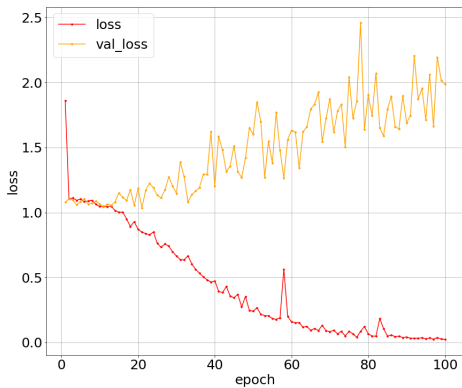


図 2: 損失率の変化

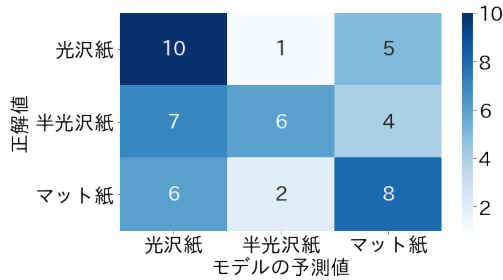


図 3: 混同行列

4.3 実験結果

図 1, 2 に結果を示す. 縦軸の loss と acc は学習データの誤差と認識精度, val_loss と val_acc は評価データの誤差と認識精度であり, 横軸は epoch 数である. epoch 数が増えるごとに学習データの認識精度が 1.0 近くになっていることから, 学習ができていることがわかる.

表 4: テストデータでの考察

	特徴として注視した箇所
光沢紙	<ul style="list-style-type: none"> ・ 光沢感 ・ シャープで明確である ・ 焦点が合っている
半光沢紙	<ul style="list-style-type: none"> ・ 彩度が高い部分 ・ シャドウ部分 ・ 焦点が合っていない背景
マット紙	<ul style="list-style-type: none"> ・ 彩度が高い部分 ・ ハイライト部分

4.4 考察

図 2 より, loss は低くなっていくが val_loss は上昇傾向にある. これは, 学習データセット数が足りず過学習が生じていると考える. 学習に使用していないテストデータ (各用紙 16~17 枚) を用いて識別をしたところ, 認識精度は 47% となった. 図 3 に混同行列を用いた結果を示す. 混同行列の結果より, 光沢紙は半数以上分類できているのに対し, 半光沢紙とマット紙が分類できていないことがわかった. 表 4 に特徴量可視化をした各用紙の特徴を示す. 光沢紙は光沢感やシャープで明確である箇所が注視されていることがわかった. また, 光沢紙と予測した画像すべてにおいて焦点が合う箇所が注視されており, 背景などの焦点が合っていない箇所は注視されていない. これは焦点が合っていない箇所はシャープさに欠けているため注視されていないと考える.

半光沢紙はテストデータの正答枚数が半数以下となり 3 種類で一番少ない結果であった. 半光沢紙の特徴量として, 彩度が高い箇所やシャドウとなっている箇所などが挙げられるが, すべての写真に光沢紙の特徴量であるシャープで明確である箇所が含まれているため, 誤分類が多くなったと考えられる.

マット紙は彩度が高い箇所やハイライトが強い箇所が注視されていることがわかった. ハイライトが強い特徴は他の 2 種ではみられなかった箇所でありデータの半数が正解する結果につながったと考察できる. マット紙も半光沢紙と同様にシャープの特徴が含まれており光沢紙の誤分類が増えたと考えられる. これらの事より, 光沢紙の特徴量が他の用紙で誤分類される原因となっている可能性が高く, シャープで明確である以外の特徴量を見つけるためにデータセットの改善や枚数の増加が必要であると考察できる.

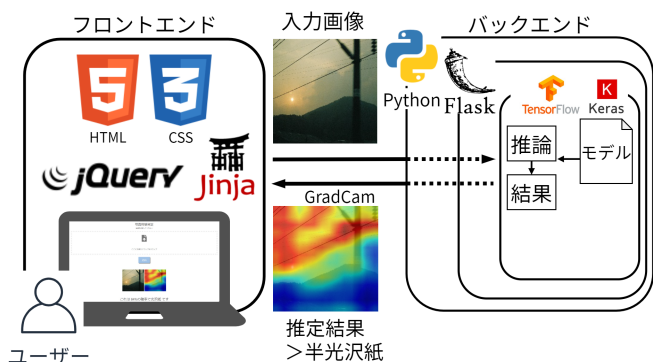


図 4: システム構成図

5 応用

実験で作成したモデルを用いて任意のユーザ画像を元に推論結果と分類寄与箇所を出力する Web アプリを作成した。バックエンドに Python の Web アプリケーションフレームワークである Flask, フロントエンドに HTML, CSS, jQuery, Jinja2 を使用した。図 4 にシステム構成図を示す。この Web アプリはサイトにアクセス後、任意の画像を選択しサーバーへ送信する。サーバーでの推論が終了後ユーザー画面に推論結果と分類寄与箇所の画像が表示される仕組みである。寄与箇所のヒートマップは赤色, 黄色, 青色と変化していく。赤色が分類に大きく注視した箇所であり, 青色が分類にあまり注視しなかった箇所である。

6 まとめと今後の課題

本研究では写真に調和する写真用紙を推定・提示するシステムを提案する事を目的とし, CNN を用いた画像分類の実験をし, Web アプリを作成した。実験の結果, 識別が可能である事が示唆された。用紙種別においてマット紙は他の 2 種に比べ特徴が出やすいが, 光沢紙と半光沢紙は特徴が同じ箇所があり分類が難しいことがわかった。用紙種別を 3 種類のグループ化しない製品別での実験結果とも合わせると, 学習率などのハイパーパラメータの調整よりデータの前処理の部分が識別率に影響することがわかっている。今後の課題として, データセットを改善し精度の高いモデル作成が必要である。また, 創作者が用紙を選んだ理由と CNN が推定した結果が一致するか確認し精度を向上させることを目標とする。

謝辞

データセットの制作に対し, データを用いた実験を快諾いただいたダイヤモンド (株) (ピクトリコビジネスユニット) の皆様に感謝申し上げます。

参考文献

- [1] 北野賀久, 江前敏晴. オフセット印刷と電子写真における画像構造の比較. 日本画像学会誌, Vol. 49, No. 1, pp. 26–33, 2010.
- [2] J. Deng, W. Dong, R. Socher, L. Li, Kai Li, and Li Fei-Fei. Imagenet: A large-scale hierarchical image database. In *2009 IEEE Conference on Computer Vision and Pattern Recognition*, pp. 248–255, June 2009.
- [3] Karen Simonyan and Andrew Zisserman. Very deep convolutional networks for large-scale image recognition. In *International Conference on Learning Representations*, 2015.
- [4] 中山英樹. 深層畳み込みニューラルネットによる画像特徴抽出と転移学習. 電子情報通信学会技術研究報告 (= IEICE technical report) : 信学技報, Vol. 115, No. 146, pp. 55–59, 2015.
- [5] Ramprasaath R. Selvaraju, Michael Cogswell, Abhishek Das, Ramakrishna Vedantam, Devi Parikh, and Dhruv Batra. Grad-cam: Visual explanations from deep networks via gradient-based localization. *International Journal of Computer Vision*, Vol. 128, No. 2, p. 336–359, Oct 2019.