

Photograph Quality Prediction using Convolutional Neural Networks

Ryoma Okuno, Yasuaki Ito and Koji Nakano

Graduate School of Advanced Science and Engineering, Hiroshima University

Kagamiyama 1-4-1, Higashi-hiroshima, 739-8527, JAPAN

Abstract—In this paper, we propose a machine learning method to evaluate the quality of photographs with high accuracy. Specifically, the assessment of photographs is performed as a problem of classifying whether the input photograph was taken by an amateur or an expert photographer. The proposed network model utilizes ConvNext, a convolutional neural network, and provides histograms as auxiliary input as color distribution information to improve the accuracy. The experimental results show that the proposed model achieves a test accuracy of 92.5%.

Index Terms—photograph quality prediction, CNN, ConvNext, machine learning

I. はじめに

近年, Instagram などの写真を投稿する SNS が普及したことで, 撮影した写真を公開する機会が増え, さらにスマートフォンに搭載されているカメラの性能が飛躍的に向上しているため, 多くの人がより良い写真を撮ることに注目している。しかし, 一般人には撮影した写真の評価は簡単でないため, SNS に投稿する際にどの写真を公開するべきか選択することが難しいとされる。そこで本研究では機械学習を用いて写真の品質を評価する手法を提案する。具体的には専門家が撮影した写真とアマチュアが撮影した写真を分類する分類問題として写真の品質評価を行う。

本研究では写真全体の構図と写真の色情報に注目して品質評価を行うように学習した分類モデルを提案する。提案手法のモデルでは畳込みニューラルネットワークにヒストグラムを用いた色分布情報を追加することで精度の向上を行った。本研究では専門家の写真には Unsplash [1] から収集した写真を使用し, アマチュアの写真には Instagram や著者らが撮影した写真を使用した。そのため, 本研究で作成した分類モデルは厳密にいうと, 入力された写真が Unsplash の写真か否かを分類する分類モデルである。しかし, Unsplash に写真を掲載するには, 審査を受ける必要があり, 一定以上の技術がないと掲載することができないため, 誰でも気軽に投稿できる Instagram と比べ品質に十分な差があると考えられる。そのため, 本研究で作成した分類モデルは入力した写真を専門家が撮影したものなのか, アマチュアが撮影したものなのかに分類することができる分類モデルであると考えられる。この分類問題に対して, 実験の結果テスト精度 92.5% で専門家とアマチュアの撮影した写真を分類できることを示した。

本論文の構成について説明を行う。本論文は 7 節で構成されており, 第 2 節では機械学習を用いた写真の品質評価手法についての先行研究についての紹介, 第 3 節では使用したデータセットについて, 第 4 節では提案手法, 第 5 節では実

験結果, 第 6 節では作成した AI の分析, 最後に第 7 節では結論を述べる。

II. 先行研究

近年, 機械学習が発展したことで写真などの芸術作品からより詳細な特徴を抽出することが可能となり, 抽出した特徴を用いて品質評価を自動で行う研究がさまざまな視点から行われている。

最初に写真の自動品質評価に関する既存研究について簡単に説明を行う。写真の自動品質評価についての既存研究の中には, 単純に高品質写真と低品質写真の 2 クラス分類を行うモデルを作成することで品質評価を自動で行う研究 [6], [7] だけでなく, 入力された写真を評価するまえに, 内容に応じて分類を行い, それぞれの分野に応じて異なるネットワークモデルを用いて写真の品質評価を行う研究 [14] や, 広大な範囲を撮影した写真と一点に集中した写真では撮影を行う際のルールが異なるため, あらかじめ分類を行い, 写真の注目しているところを事前に特定し品質評価を行う研究 [8] のように, 品質評価前に写真のスタイルに応じて分類を行う種類の研究が存在する。これ以外にも, シャムニューラルネットワークを利用することで, 高品質写真と低品質写真の 2 クラス分類だけでなく, 一般的な写真と比較した際の品質や自身が過去に撮影した写真と比較した際の品質を同時に分析するネットワークについての研究 [9]。写真の色彩情報により焦点を当てて品質評価を行うために, 1 枚の写真をかなり細かく局所的に分割することで一つ一つの局所領域の色彩情報を単調なものとし, それぞれの領域を Moon-Spencer と呼ばれる色彩調和論で用いられている指標を用いて色彩調和スコアを計測することで写真の品質を評価している研究 [10]。局所的に分割することは汎用性を下げる恐れがあるため, 深くて広い CNN を用いることでできるだけ全体的な特徴量を入力から抽出する研究 [11] が存在する。

また, 以上のような写真の品質評価以外にも, 機械学習を用いてユーザーがより良い写真を撮影するための手助けをする研究 [5]。どのような写真なのか形容詞を用いて分類する研究 [12] や写真の状況を言語で予測する研究 [13] などさまざまな先行研究が存在する。

本研究では, 入力に写真だけではなく色彩情報のヒストグラムを入力することで, 予測モデルに写真全体の構図を認識させ, 色彩情報を重視し品質評価を行うことを目標に研究を行っている。

III. データセット

本研究では, 収集した専門家の写真とアマチュアの写真で構成されているデータセットを用いて実験を行う。それぞれ

の写真の収集方法については次節で説明を行う。

A. 専門家写真

専門家写真は Unsplash [1] から収集した。Unsplash では写真を投稿する際に一定以上の技術がないと投稿できず、構図などの問題で投稿を拒否される場合はアドバイスをしてもらうことができる。また、著作権もかなりフリーで、310 億回以上ダウンロードされているオープンなフォトサイトなため、多くのプロの写真家が写真を投稿しており、現在では 4 万人以上在籍している。

今回のデータセットでは 10000 枚の写真を Unsplash から収集し、訓練には 9000 枚、テストには 1000 枚重複しないように用いる。また、訓練時には写真を水平方向に反転させることで写真の水増しを行ったため、18000 枚の専門家の写真を用いて学習を行う。Fig.1 は今回使用した専門家写真の一例を示す。



Fig. 1. Example of professional photo dataset

B. アマチュア写真

アマチュアが撮影した写真は Instagram から収集したものと、著者らが自ら撮影したもので合計 10000 枚になるように構成した。専門家写真と同様に、訓練には 9000 枚、テストには 1000 枚重複しないように用い、訓練時には写真を水平方向に反転させることで写真の水増しを行ったため、専門家写真と同様に 18000 枚のアマチュア写真を用いて学習を行う。Fig.2 は今回使用したアマチュアが撮影した写真の一例を示す。

IV. 提案手法

この節では提案手法に用いられているネットワークモデルについて説明を行う。本研究では、ネットワークモデルの基盤に ResNet50 [2], または ConvNeXt [4] を用いる。ResNet50 や ConvNeXt は下記の Fig.3 に示すブロックを用いることで構成されている。ブロックを用いたより詳細な ResNet50 と ConvNeXt のネットワークモデルは Fig.4 と Fig.5 にそれぞれ示す。

ResNet は ResNet50 や ResNet152 などのように層を深くするように拡張していたが、ConvNeXt ではただ単純に層を深くするだけでなく、それぞれのチャンネル数を変更することで TABLE I に示すように 5 つのネットワークモデルに拡

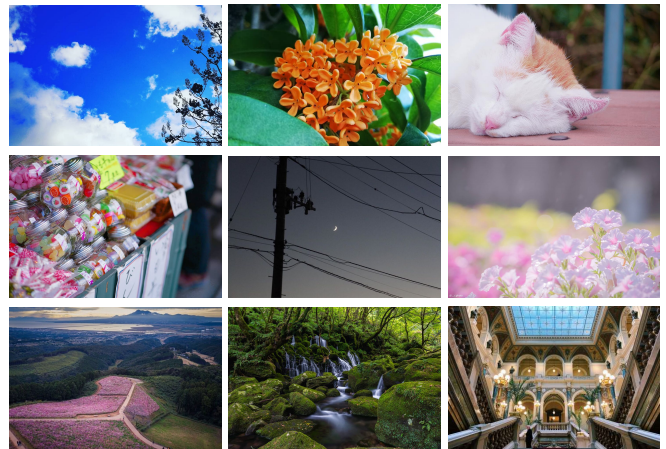


Fig. 2. Example of beginner photo dataset

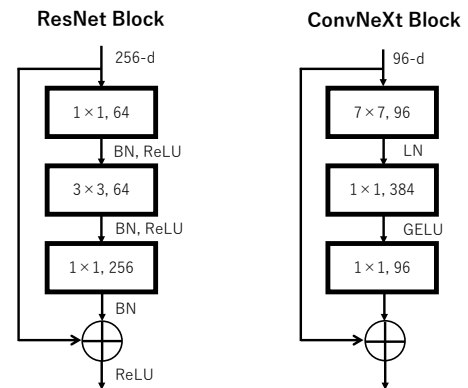


Fig. 3. Network Block

張している。本研究ではそれぞれの ConvNeXt のタイプで実験を行ったが、ConvNeXt-Tiny が最も精度が高かったため、実験を行う際は ConvNeXt の中でも Tiny サイズのものを使用する。

TABLE I
NETWORK STRUCTURE OF CONVNEXT

Type	Number of Blocks	Number of channels in each Block
ConvNeXt-Tiny	(3, 3, 9, 3)	(96, 192, 384, 768)
ConvNeXt-Small	(3, 3, 27, 3)	(96, 192, 384, 768)
ConvNeXt-Base	(3, 3, 27, 3)	(128, 256, 512, 1024)
ConvNeXt-Large	(3, 3, 27, 3)	(192, 384, 768, 1536)
ConvNeXt-XLarge	(3, 3, 27, 3)	(256, 512, 1024, 2048)

本研究で提案するネットワークでは、入力された写真の色情報に特に注目して学習させる。そのため、入力写真全体を分析するネットワークと入力写真の色情報のみを分析するネットワークの二つを用いて学習を行い、Fig.6 に示すように、それぞれのネットワークの出力結果をまとめて全結合層に入力することで予測値を出力する。この際、出力チャンネルを 2 にすることで予測結果を二変数で出力し、ソフトマックスを用いて分類結果を出力する。入力写真全体を分析するネットワークには ResNet50 や ConvNeXt を用い、色情

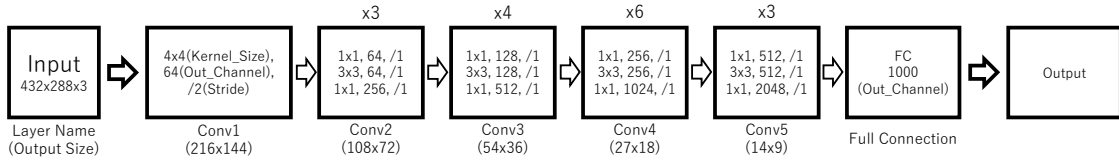


Fig. 4. ResNet50 Model

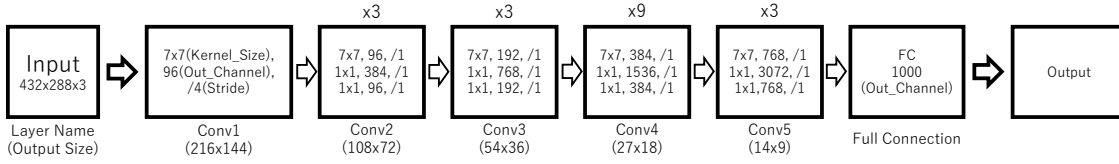


Fig. 5. ConvNeXt-Tiny Model

報を分析するネットワークには全結合層 3 層を用いて分析を行う。

トグラムでは bin の本数は 16 本だが、本研究で作成したヒストグラムの bin の本数は 256 である。

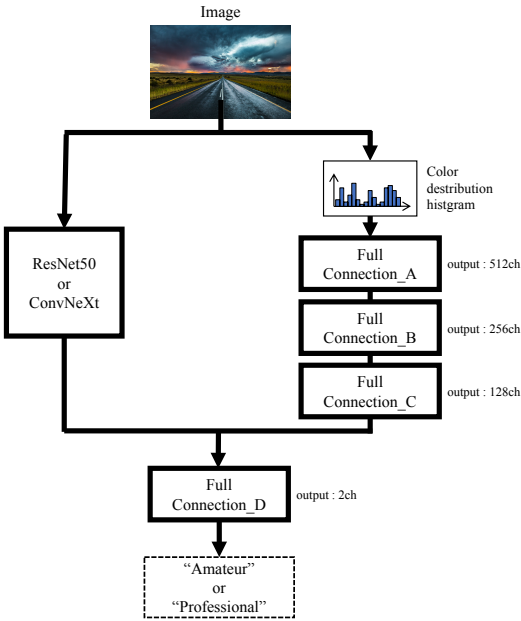


Fig. 6. Network Overview

本研究では、色情報を抽出する際に RGB の色調空間を用いる。ネットワークの全体像は Fig.6 に示す。

A. ヒストグラムの作成方法

本研究で使用する色分布情報ヒストグラムの作成方法について説明を行う。本研究では、Fig.7 に記載されているとおり、色分布情報としてヒストグラムと累積ヒストグラムの 2 通りの方法で作成し、利用している。ヒストグラムは入力された写真をもとに累積和を使用せずヒストグラムを作成し、ヒストグラムの最大度数が 1 になるように正規化を行うことで作成する。累積ヒストグラムは、累積和を使用してヒストグラムを作成し、前述のヒストグラムと同様に最大度数が 1 になるように正規化を行うことで作成する。Fig.7 のヒス

V. 実験

本節では ResNet [2] や ConvNeXt [4] を使った実験結果について紹介する。本実験では最適化アルゴリズムには Adam を、損失関数には Cross Entropy を使用し、エポック数は 50 である。

A. 分類モデルの実験結果

TABLE II より、本研究の最高精度は 92.5% であることがわかる。また、ネットワークの基盤に ResNet50 よりも ConvNeXt を使用した方が精度が高く、Histogram を追加で入力することでさらに精度が向上する傾向があることがわかった。この結果より、写真から特徴量を抽出する際は ResNet50 よりも ConvNeXt の方が効果的であり、写真の分類を行う上で色情報を追加で入力することは有効的であることがわかった。

TABLE II
EXPERIMENTAL RESULT

Network	Color distribution	Learning Rate	Test Accuracy (%)
ResNet50	—	1e-05	88.2
ResNet50	Histogram	1e-05	89.5
ResNet50	Cumulative Histogram	1e-05	88.9
ConvNeXt	—	1e-05	91.7
ConvNeXt	Histogram	1e-05	91.6
ConvNeXt	Cumulative Histogram	5e-06	92.5

次に、Fig.8 に示す正解ラベルと分類ラベルの混同行列を用いて実験結果の分析を行う。

最初に適合率について分析を行う。Fig.8 より、専門家であると分類した結果の適合率が約 94.1%、アマチュアであると分類した結果の適合率は約 89.3% である。この結果から、分類した結果の信頼性はかなり高いことがわかる。特に、専門家であると分類された時の信頼性が特に高いことがわかる。

次に再現率について分析を行う。Fig.8 より、専門家であると分類した結果の再現率は 90.7%、アマチュアであると分類した結果の再現率は 94.3% である。この結果から、再現率が高いため、正確に分類できていることがわかる。特にアマ

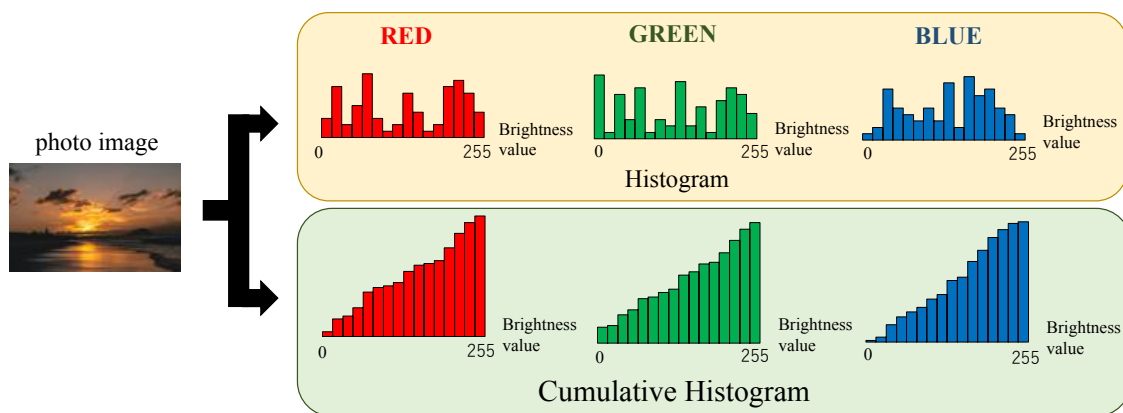


Fig. 7. Color distribution information indicated by histograms and cumulative histograms

チュアの写真を専門家の写真だと誤認する確率が5.7%とかなり低いことがわかった。

最後にF値についての分析を行う。前述したとおり専門家である適合率と再現率はそれぞれ約94.1%, 90.7%である。これより、専門家のF値は約0.924。同様に、アマチュアのF値は約0.926となる。専門家、アマチュアのどちらも適合率と再現率がそれぞれ高いため、F値がかなり1に近い値となり、高精度で分類できていることがわかる。

B. 分類結果についての分析

この節では作成した分類モデルに対して実際に写真を入力し、どのように評価されているか確認する。Fig.8ではアマチュア写真1000枚、専門家写真1000枚の合計2000枚の写真を本実験で作成した学習モデルに入力した際の結果の一例を示す。図に示すように、横軸が正解ラベルを示し、縦軸が学習済みモデルが分類した結果を示す。図の説明として右上のエリアについて説明を行う。右上のエリアの写真は、正解ラベルがアマチュアであり、分類結果は専門家である。下の“57 images”というのは、1000枚のアマチュア写真の中で専門家だと分類された写真が57枚あることを示す。この結果を見ると、分類モデルがアマチュアであると判断した写真の色情報は単調なものが多く、専門家と判断した写真の色情報はカラフルであることが確認できる。また、専門家と判断している写真の中には、見せたい部分のみ明るくし、他の部分を暗くするトンネル構図や、背景をぼかすなどの技術が活用されていることがわかるため、色情報だけでなく、ある程度の構図を理解していることが期待できる。

VI. まとめ

本論文では、自身で収集した画像を用いて、専門家の写真なのか、アマチュアの写真なのかを分類することを目的に実験を行った。本研究では、専門家の撮影した写真をUnsplash [1] と呼ばれる、一定以上の技術がないと投稿することができないフォトサイトから収集し、アマチュアの撮影した写真をInstagramや自身で過去に撮影した写真から収集することでデータセットとして用いた。ConvNeXtを用いることで専門家の写真とアマチュアの写真を高精度に分類することが可能なモデルを作成することができた。そして、色情報のヒストグラムを学習に加えることでより分類精度を向上させる

ことができた。本研究で作成したモデルは厳密にはUnsplashで投稿されるレベルの写真とInstagramで投稿されるレベルの写真を分類することができる分類モデルである。しかし、UnsplashはInstagramと比べ、投稿する際に審査が必要となるため、Unsplashの写真は専門性が高いと考えられる。従って、本研究で作成したモデルは写真をプロレベルとアマチュアレベルを分類するモデルであると考えられる。

REFERENCES

- [1] <https://unsplash.com/>.
- [2] Kaiming He, Xiangyu Zhang, Shaoqing Ren, Jian Sun, “Deep Residual Learning for Image Recognition”, arXiv:1512.03385v1, 2015
- [3] Karen Simonyan, Andrew Zisserman, “Very Deep Convolutional Networks for Large-Scale Image Recognition”, arXiv:1409.1556v6, 2015.
- [4] Zhuang Liu, Hanzi Mao, Chao-Yuan Wu, Christoph Feichtenhofer, Trevor Darrell, Saining Xie, “A ConvNet for the 2020s”, arXiv:2201.03545v2, 2022
- [5] 田中直人, 高島健太郎, 西本一志, 機械学習技術を応用した写真撮影における個性発揮支援システムの提案, 情報処理学会研究報告. HCI, 研究報告ヒューマンコンピュータインタラクション, pages 1-8, 2018.
- [6] Fei Gao, Ziyun Li, Jun Yu, Junze Yu, Qingming Huang, Qi Tian, “Style-adaptive photo aesthetic rating via convolutional neural networks and multi-task learning”, Neurocomputing, pages 247-254, 2020.
- [7] Yan Ke, Xiaou Tang, Feng Jing, “The Design of High-Level Features for Photo Quality Assessment”, IEEE, 2006.
- [8] Quyet-Tien Le, Patricia Ladret, Huu-Tuan Nguyen, Alice Caplier, “Image Aesthetic Assessment Based on Image Classification and Region Segmentation”, MDPI Journal of Imaging, 2021.
- [9] Jun-Tae Lee, Chang-Su Kim, “Image Aesthetic Assessment Based on Pairwise Comparison -A Unified Approach to Score Regression, Binary Classification, and Personalization”, IEEE, 2019.
- [10] Masashi Nishiyama, Takahiro Okabe, Imari Sato, Yoichi Sato, “Aesthetic Quality Classification of Photographs Based on Color Harmony”, IEEE, 2011.
- [11] Yunlan Tan, Pengjie Tang, Yimin Zhou, Wenlang Luo, Yongping Kang, Guangyao Li, “Photograph aesthetical evaluation and classification with deep convolutional neural networks”, Neurocomputing, pages 165-175, 2017.
- [12] Yaowen Wu, Christian Bauckhage, “The Good, the Bad, and the Ugly: Predicting Aesthetic Image Labels”, ICPR, 2010.
- [13] Sagnik Dhar, Vicente Ordonez, Tamara L Berg, “High Level Describable Attributes for Predicting Aesthetics and Interestingness”, IEEE, 2011
- [14] Hanghang Tong, Mingjing Li, Hong-Jiang Zhang, Jingrui He, “Classification of digital photos taken by photographers or home users”, In Advances in Multimedia Information Processing, pages 198-205, 2004.

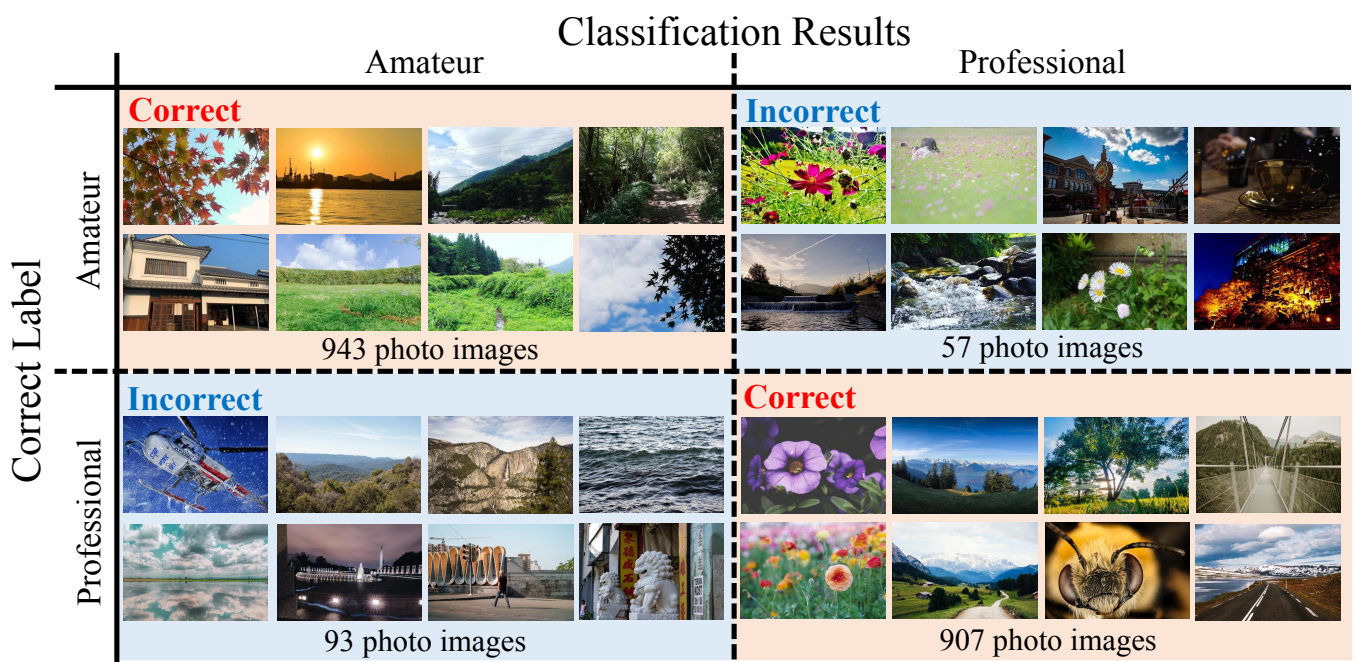


Fig. 8. An example of classification results