ISSN 2186-5647

5-7

CNN を用いた、風景画像の評価構造の抽出に関する研究
-SNS における注目度の違いに着目して -

日大生産工（院）　○鈴木　俊策
日大生産工　岩田　伸一郎

1. はじめに

Instagram は、多くの人々が利用している人気の SNS アプリケーションであり、2017 年の流行語大賞に選ばれた「インスタ映え」という言葉のように、写真が重要視されているアプリケーションである。また、近年、写真に映える場所である「インスタ映えスポット」が人気を集め、Instagram の投稿をきっかけとなり、多くの人々が、その場所を訪れることで観光のきっかけとなる事例が増えている [1]。

そこで本稿では、写真を重要視する Instagram に投稿された風景画像に着目し、投稿された画像の注目度をもとに分析を行う。分析手法は、現在、画像認識の分野で人と同じかそれ以上の認識制度を達成している畳み込みニューラルネットワーク [2]（以下、CNN）を用いて機械学習を行い、学習の過程を抽出することで、Instagram で注目を集めやすい風景画像の色の分布や特徴を抽出することを目的とする。

2. 関連研究

CNN は、画像を受け取る入力層から、入力画像の特徴を抽出する畳み込み層と、プーリング層を何度か繰り返すことで特徴を抽出していく。その後、何層かの全結合層を経て出力に至る構造を基本としている。この構造は、LeCun らにより、人間の視覚野が物体を認識する際の構造を模倣した分析手法として提案された [3]。その為、CNN の有用性は、幅広い問題に及ぶ事もわかっており、特に VGG-16 [注 1)] や、ResNet [注 2)] のような画像認識の分野で非常に高い性能を示している [4]。

CNN を用いた風景画像の研究では、 神戸らが、服飾・風景画像を入力値に、印象評価実験により 4 段階評価された印象語を出力値として付与し、画像の印象推定を定量的に行なっている [5]。
また、飛谷らは、プロダクトデザインの分野において、腕時計のデータセットを入力値に、出力値に被験者実験によりえた 7 段階の印象評価値を取得し、学習に用いている [6]。これらの研究から CNN は、単純な画像認識だけでなく、感性情報を元にした画像認識も可能であることが示されている。

これらのことから、本手法において、SNS 上の投稿においての他のユーザーからの注目度を求める事ができるエンゲージメント率 [注 3)] を用いて分析することも有用だと考えられる。

$$\text{エンゲージメント率 (\%)} = \frac{\text{いいね数}}{\text{フォロワー数}} \quad \cdots(1)$$

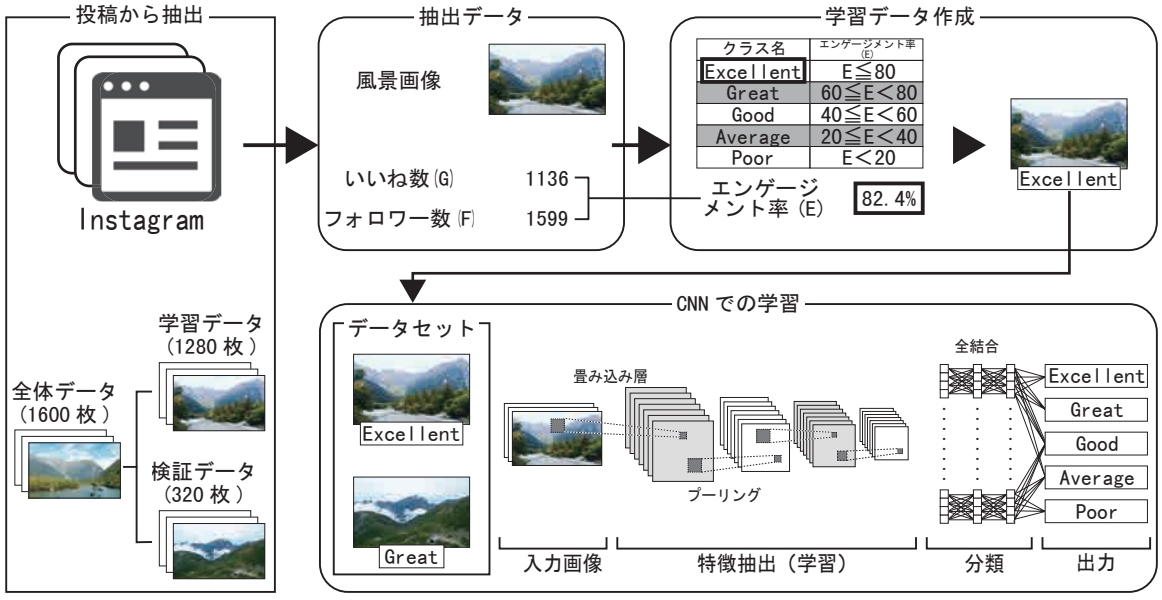

図 1　学習データセット作成と CNN 実装までのフロー

Research on the relationship between train use and staying places and walking smartphones.
-Focusing on the difference in attention in SNS-

Syunsaku SUZUKI, Shinichiro IWATA

クラス名	Excellent	Great	Good	Average	Poor
エンゲージメント率の割合	80% 以上	60～80%	40～60%	20～40%	20% 未満
風景画像例（エンゲージメント率）	 83.6%	 73.2%	 45.5%	 39.5%	 18.3%
	 82.4%	 68.3%	 50.6%	 20.5%	 7.4%

図 2　データセットの代表例

3. 研究方法

3.1 対象データセット

本手法では、図 1 のように Instagram から、ユーザーが投稿した風景画像（1600 枚収集。1280 枚を学習データ、320 枚を検証データとする）を取得し、CNN の入力データとする。さらに、いいね数、フォロワー数も収集する（いいね数、フォロワー数が 10 未満のものは対象データとしない）。いいね数、フォロワー数は、式(1)からエンゲージメント率を算出し、値ごとに 5 つのクラスに分類する。5 つのクラスに分類された値を学習の出力データとする。そうすることで入力された風景画像から、注目度を元にした CNN の学習データを作成する。

表 1　構築した CNN の構造

層	入力サイズ	出力サイズ	カーネルサイズ	スライド幅
畳み込み層 1	250×250×3	250×250×16	3×3	1
最大プーリング層 1	250×250×16	125×125×16	2×2	2
畳み込み層 2	125×125×16	125×125×32	3×3	1
最大プーリング層 2	125×125×32	62×62×32	2×2	2
畳み込み層 3	62×62×32	60×60×64	3×3	1
最大プーリング層 3	60×60×64	30×30×64	2×2	2
畳み込み層 4	30×30×64	30×30×64	3×3	1
最大プーリング層 4	30×30×64	15×15×64	2×2	2
全結合層 1	15×15×64	1×1×14400	-	-
全結合層 2	1×1×14400	1×1×14400	-	-
全結合層 3	1×1×14400	1×1×5	-	-

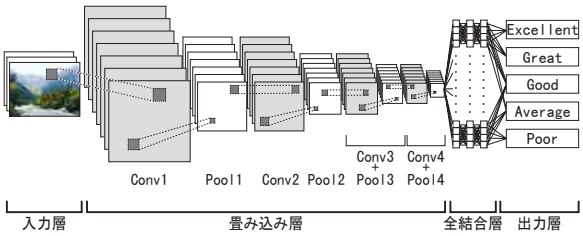

図 3　構築した CNN の構造図

3.2 CNN の構築

本手法で構築した CNN は、畳み込み層 13 層、全結合層 6 層の全 19 層の CNN とした。CNN の構造を図 3、表 1 にそれぞれ示す。

CNN の学習において、バッチサイズは 64、エポック数は、20 とする。活性化関数は Rectified Liner Unit(Relu) 関数[注 4)]（以下、Relu 関数）を使用し、全結合層でも Relu 関数を使用する。出力層では sigmoid 関数を使用し、損失関数[注 5)]による誤差を最小化するように CNN の重みを更新することで学習の精度を高めていった。

3.3 印象推定結果

学習した CNN の結果、図 4、表 2 において、学習全体の正答率は 81.5% となった。全学習データは 93.3% の正答率、損失率も 9% となった。しかし検証データが正答率 32.5%、損失率 151.8% となったことにより過学習を起こしている事がわかる。

表 2　CNN の正答率と損失値

CNN の学習結果	全体データ	学習データ	検証データ
正答率 (%)	81.5	93.8	32.5
損失率 (%)	40.9	9.8	151.8

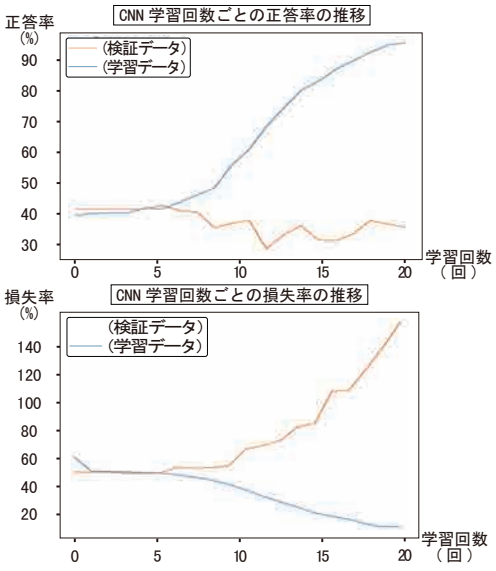

図 4　CNN の学習の過程

3.4 推定結果の分析

CNN は高精度な画像認識手法であるが、どのような観点で認識されているのかは、理論的に解明されていなかったが、近年、その判断根拠を可視化する研究[7]が行われている。また、鈴木らが、CNN の推論過程を可視化する研究[8]を行い、CNN の判断根拠を分析することが可能になりつつある。

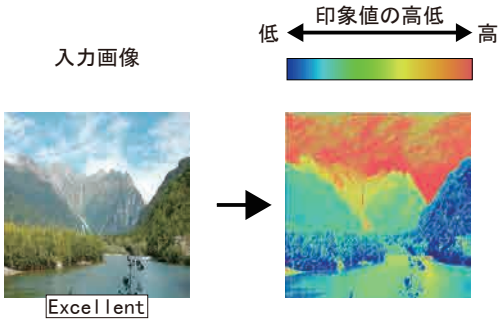

図 5　Grad-CAM による可視化の例

本稿では、図 5 のように Grad-CAM[注6)] によるヒートマップ画像を作成し、判断根拠の可視化を行う。入力画像と出力値に対して学習過程において重要な箇所が赤く表示され、低いところは青く表示される。

4. 考察

4.1 CNN の学習モデル作成において

本手法は、CNN を用いて SNS に投稿された注目を集めやすい風景画像の特徴を抽出しようと試みた。しかし構築した CNN は、過学習を起こし、学習結果としては改善が必要である。過学習を起こした原因としては、

1) 学習データの不足

2) 学習データセットが高次元により分類が困難

の二つが挙げられる。1) はデータ数を増やせば良いが、2) は、fine-tuning を行い学習済みモデルから新たな学習データを作成すれば改善が可能である。

4.2 Grad-CAM における考察

図 6 のヒートマップ画像から注目度の高い所は、明度の高い空や海などが赤く表示され、森などの明度の低い所が青く表示された。これは、CNN の学習が正しく行われていないといえる。CNN のプーリング層では、MaxPooling を行なっている。これは、特徴を値として抽出した中で一番大きな値を抽出する手法である為、明度が高いほど赤く表示されたと考えられる。Pooling 層には、特徴の平均値を抽出する AveragePooling もある為、今後は併用しながらパラメータを決定していく。

5. まとめ

本手法で構築した CNN は正答率 81.5% と、まずまずの結果となったが、過学習を起こしていた事、また

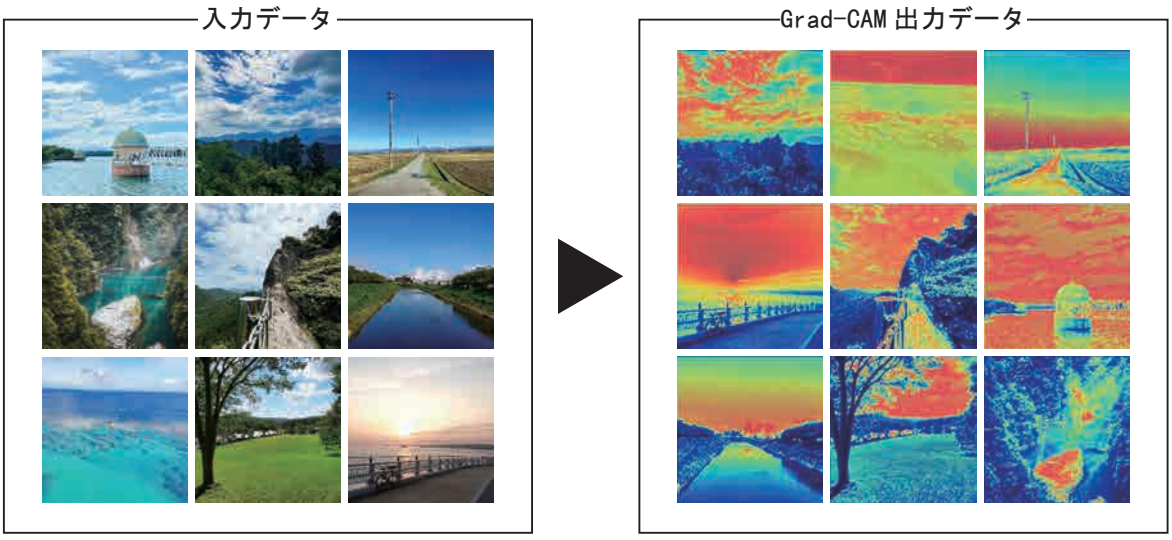

図 6　入力データとヒートマップ画像の例

Grad-CAM を用いた分析においても有益な結果とはならなかった。今後、学習モデルをさらに高め、分析手法を確立するために、

1) 過学習を抑制する為、各種パラメーターの改善、学習データを増やす。
2) 風景画像の学習済みモデルを用いて fine-tuning[注7)] を行い学習精度を高める。
3) 分析手法の改善として、本手法では、鈴木らが提案した CNN の推論課程を可視化する手法は、行なっていない。今後は、分析手法を決定するために、実装して分析手法を決定する。

以上、三点を踏まえ今後の研究を進めていこうと考える。

注釈

注1) 「ImageNet」と呼ばれる莫大な画像データセットで学習された、畳み込み層 13 層、全結合層 3 層の全 16 層のニューラルネットワーク。入力された画像を 1000 のクラスに分類が可能。

注2) 全 152 層のニューラルネットワークであり、画像認識の分野で非常に高い性能を示している。

注3) ある投稿に対してどれくらいのエンゲージ（反応：いいね、クリック、シェア）があったかを測る指標。各 SNS により算出方法が異なり、Instagram では式 1) により求める事ができる。

注4) 関数への入力値が 0 以下の場合に出力値が常に 0、入力値が 0 より上の場合には出力値が入力値と同じ値となる関数。

注5) CNN における損失関数とは「正解値」とモデルにより出力された「予測値」とのズレの大きさを計算する為の関数。

注6) CNN で学習された予測値に対する勾配を重みづけする事で、画像の重要なピクセルを可視化する技術。

注7) 学習済みの CNN の一部、もしくは全ての層の重みを微調整することで、学習済みモデルを元にして新たな学習モデルを作成する手法

参考文献

1) JTB 総合研究所,「スマートフォンの利用と旅行消費に関する調査」(2019), https://www.tourism.jp/wp/wp-content/uploads/2019/11/smartphone-travel-consumption.pdf（閲覧日：2021,10,08)
2) He,Kaiming:「Delving Deep into Rectifiers:Surpassing Human-Level Performance on Imagenet Classification」.ICCV, 2015.
3) Y.Lecun:「Back-propagation applied to handwritten zip code recognition」.Neural Computation, Vol.1, pp.541-551, 1989.
4) 岡谷貴之:「画像認識のための深層学習の研究動向」. 人工知能，第 31 巻，第 2 号，pp.169-179, 2016
5) 神戸瑞樹：「CNN を用いた服飾・風景画像に対する印象の推定」. 情報処理学会・FIT 情報科学技術フォーラム，第 18 回，pp.93-96, 2019
6) 飛谷謙介:「スタイル特徴を利用した DNN による印象推定に寄与する画像領域の可視化」. 日本工業出版・画像ラボ，第 31 巻，pp.38-44, 2020
7) M.D.Zeiler:「Visualizing and Understanding Convolutional Netwoks」.in Proceedings of the European Conference on Computer Vision, pp.818-833 2014.
8) 鈴木航:「深層学習における推論過程の可視化の検討」. 情報処理学会，第 80 回，pp.195-196, 2018