

料理写真の構図決定補助システムにおける構図評価手法の検討

大木 慧矢[†]
Seiya Ohkiマッキン ケネス ジェームス[†]
Kenneth J Mackin永井 保夫[†]
Yasuo Nagai東京情報大学 総合情報学科[†]

1 はじめに

近年スマートフォンの発展・普及により写真撮影が手軽となり、SNS などへの公開が容易になってきた。また、撮影上での色合いなどは最近の HDR (High Dynamic Range) 機能の利用や写真の編集アプリにより調節が可能となるとともに、簡単に写真の編集が可能となってきたため、一層のクオリティーの高い写真が撮影できるようになってきた。このような機能やアプリを利用することにより、露出や色合いは簡単に編集可能であるのに対して、最適な構図を決めて上手く撮影することは専門家でないとおこなえないという問題があることがわかった。

我々はこのような問題を解決するために深層学習を使った魅力度推定器を利用し、魅力的な構図で撮影をするシステムを開発した[1]。しかしながら、このシステムでは、柿森ら[4]が使用した料理写真の構図の5つの要素（写真内における料理の数、料理の位置、料理の向き、料理の占有面積（写真にどの程度大きく写っているか）、撮影時のカメラの角度）のうち、「料理の向き」と「撮影時のカメラの角度」の2つの構図の要素のみでしか撮影が行えないことがわかった。

本研究では、上記の料理写真の構造の5つの要素のうち、従来システムで対応可能な「料理の向き」と「撮影時のカメラの角度」だけでなく、写真内における料理の位置と料理の占有面積の2つの要素を考慮し、撮影することができるシステムの開発を行う。そのため、本論文では深層学習を使い、「料理の位置」および「料理の占有面積」を考慮し構図を評価する判定器を提案する。

2 構図評価手法の提案

本提案では、SNS 上の料理写真に構図の要素がある程度影響があるのではないかと仮定した、この仮説を検証するために、まず、柿森ら[4]が提案した料理写真の基本構図をもとに Twitter 上の料理写真の構図と「いいね」の関係性を調べた。

手法として Twitter 上から「定食」のキーワードで 600 件の画像収集を行い、「いいね」の数を元に、上位 1~200, 200~400, 400~600 に分け、それぞれに料理写真の基本構図が何件含まれているか調べた結果、表 1 の様に順位が高いほど基本構図が多く含まれることが分かった。

表 1 基本構図が含まれる枚数

お皿の数				
	1枚	2枚以上	合計（件数）	
順位	1~200	21	16	37
	200~400	16	9	25
	400~600	9	12	21

そのため、Twitter 上から画像を取得しその画像の「いいね」を教師データとし、データセットの作成を行った。

また、データセットを判定器に学習させる際、位置、占有面積のみ学習を行いたいため、料理検出を行いデータセットの画像を料理のみが写った画像に加工し、VGG16 による学習を行い、3段階の評価を行う判定器を構築した。

次に、データセットの作成には Twitter 上から「ラーメン」の画像とそれぞれの画像の「いいね」の数を 1000 件収集し、「いいね」の数を順位ごとに並べ、順位を 0 から 1 の値に正規化を行い、正規化された順位を 3 つに分け教師データとしデータを作成した。

さらに、料理の構図を判定器に学習させる際に、料理の位置や占有面積のみを学習を行うため料理の検出を行った。料理検出には、物体の位置と種類を検出するのによく使用される YOLOv3 を使用した。ここでは、基本構図のお皿が 1 枚の場合を想定し、「ラーメン」の画像 100 枚学習を行い、図 1 のように料理検出を行った。

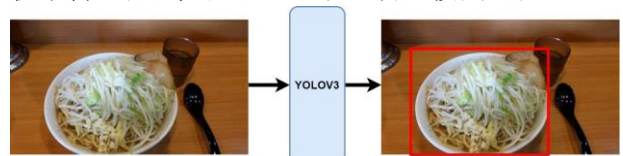


図 1 料理の検出イメージ

YOLOv3 を使った判定器に入力する際に、図 2 の料理がある領域のみを切り取った画像に加工を行った。

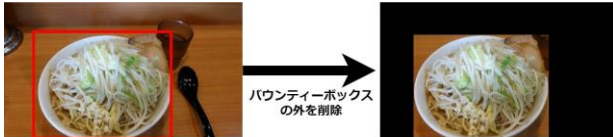


図 2 料理のみの検出を行ったイメージ

最後に、判定器の作成には前研究でも使用した VGG16 に転移学習を行った。転移学習をするにあたって作成したデータセットを 244×244 にリサイズし入力した。画像に対する教師データを入力し全結合層のみ学習を行う。重みの学習を行う全結合層の活性化関数には ReLU を使用し、出力の全結合層にはシグモイド関数を使用、最適化手法には Adam、評価関数 MAE (Mean Absolute Error) を用い、学習回数は 100 回とした。

3 評価実験

図 3 は YOLOV3 で実際にデータセットの画像から検出を行い出力した結果を示している。図 3 の A がバウンティボックスを出力した結果、図 3 の B がバウンティボックスの外側を切り取った画像をそれぞれ示している。

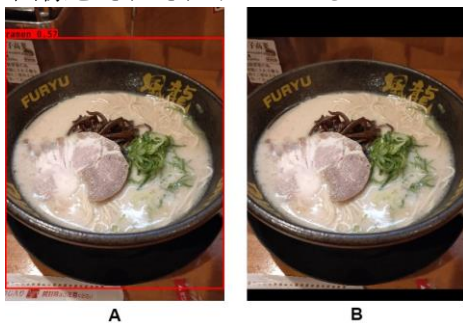


図 3 YOLOV3 でデータセットから検出を行った例

図 4 は YOLOV3 で料理がある領域のみ抽出を行ったデータセットを入力し、「いいね」の数を順位ごとに 3 段階に分類し、教師あり学習を行った正解率を示している。

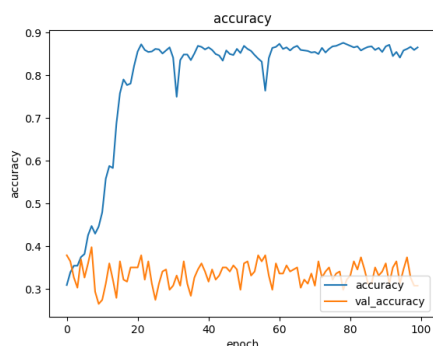


図 4 データセットを学習した正解率

料理のデータセットを作成する際に料理検出を行い料理のみの画像に加工を行った結果、100 枚という少ない画像で学習を行うことができた。また、対象が大きく映っている場合にも高い確

率で料理の検出を行うことができた。

一方、対象が小さい場合や複数の料理がある場合には、うまく検出が行えない場合があった。そのため、判定器の学習では、訓練データでの正解率は高くなったが(図4の accuracy:青線)、トレーニングデータでの正解率を高くすることができなかった(図4の val_accuracy:黄線)。

4 結論

我々が以前実施した研究では料理の料理写真の構図の要素のなかで、料理の向きと撮影時のカメラの角度のみでしか撮影が行えないことが問題となっていた。本論文では、Twitter 上の画像と画像に対する「いいね」の値を学習データとして深層学習に適用し、YOLOV3 で料理の切り抜きを行うことで料理の位置および占有面積を考慮し、判定器により構図を評価する手法を提案し、適用実験を行った。

その結果、Twitter 上のデータからデータセットの作成を行い、料理検出を行い、判定器を作成することができた。しかし、判定器の精度を向上することができなかった。

今後は YOLOV3 での検出精度を上げるため学習に使うデータの件数を増やし精度を高める必要がある。また、料理の検出できる件数を 2 枚以上検出が行えるように改良していく。さらに、Twitter 上のデータを獲得する際に料理写真以外の画像も多く取得してしいたため、料理写真のみを取得できるシステムを検討していく予定である。

参考文献

- [1] 大木 隼矢, ケネス マッキン, 永井保夫: 料理写真の魅力度推定器を使った写真撮影補助システムの検討, 第 82 回情報処理学会全国大会, pp311-312, (2020).
- [2] 佐藤 陽昇, 道満 恵介, 平山 高嗣, 井手 一郎, 川西 康友, 出口 大輔, 村瀬 洋: 畳み込みニューラルネットワークを用いた料理写真の魅力度推定, 信学技報, Vol. 117, No. 252, MVE2017-32, pp. 107-111, (2017).
- [3] Kazuma Takahashi, Keisuke Doman, Yasutomo Kawanishi, Takatsugu Hirayama, Ichiro Ide, Daisuke Deguchi, Hiroshi Murase: "Estimation of the Attractiveness of Food Photography Focusing on Main Ingredients", Proc. Of 9th Workshop on Multimedia for Cooking and Eating Activities (CEA), pp. 1-6, (2017).
- [4] 柿森 隆生, 岡部 誠, 尾内 理紀夫: おいしそうな料理写真撮影を支援するシステムの検討, 第 56 回プログラミング・シンポジウム予稿集, pp. 131-141, (2015).