

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[†] マッキンケネス ジェームス[†] 永井 保夫[†]
東京情報大学[†]

研究背景

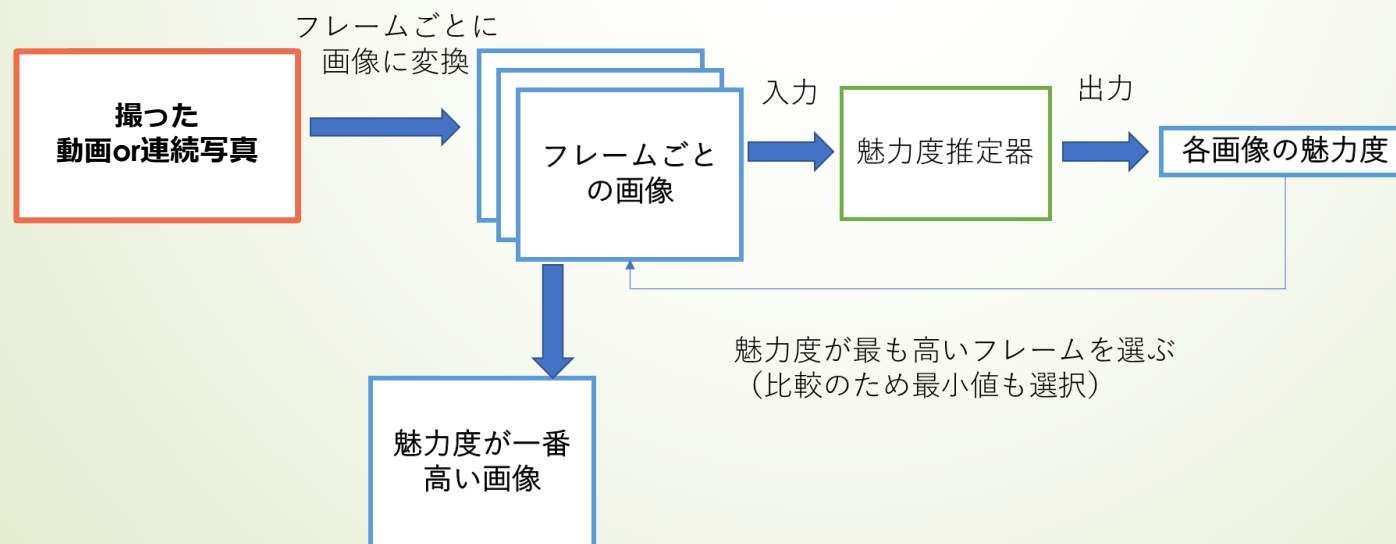
- 近年スマートフォンの発展によりSNSなどで評価を得ることができる魅力的な写真を撮りたいと言う需要が増えている
- 写真を魅力的に撮るうえで構図や色などを考慮する必要がある
- 写真を撮る上で色は写真編集アプリなどで簡単に手直しができる
- しかし非専門家が構図を考えて写真を撮るのはなかなか難しい
- 構図を考慮した魅力的な写真を撮る補助をするシステムを作れないか？

前回の提案

- 深層学習を使い、被写体の向きや角度など考慮して写真を撮影する補助することができないか
- SNSなどに上げる機会の多い料理写真を魅力的に撮る
- 魅力的な写真 = おいしそうに見える写真[2]
- 料理写真がおいしそうに見える度合いを魅力度とする
- 魅力度推定器を利用し、魅力的な撮影時の向きと角度で写真を撮る補助をするシステムの開発を行った

前回の提案

- 畳み込みニューラルネットワーク（CNN : ConvolutionNeuralNetwork）のVGG16とVGG19を使い魅力度推定器を作成
- スマートフォンから様々な方向で撮影した料理の動画と連続写真から魅力度が高い視点の写真を出力を行った



前回の研究結果

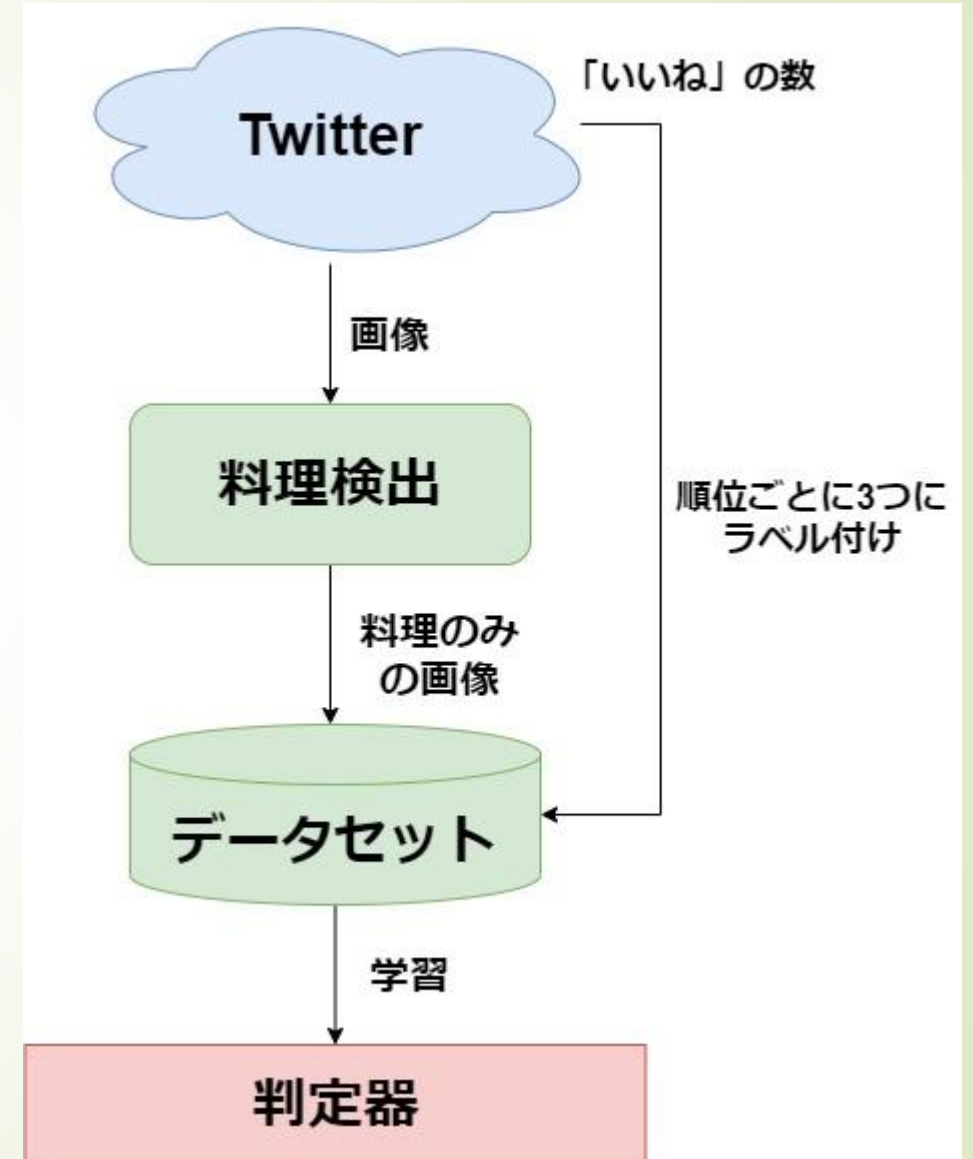
- 提案する撮影方法によりデータセットに近い（魅力度の高い）画像を選び出すことができた
 - 魅力度推定器を使用することでユーザが配置や撮る視点を考慮する必要がなく撮影が行える
- 柿森ら[4]が使用した料理写真の構図の5つの要素（写真内における料理の数、料理の位置、料理の向き、料理の占有面積(写真にどの程度大きく写っているか)、撮影時のカメラの角度)
- 「料理の向き」と「撮影時のカメラの角度」の2つの構図の要素のみでしか撮影が行えない

目的

- 料理写真の構図の5つの要素のうち写真内における料理の位置と料理の占有面積の2つの要素を考慮する
- 2つの要素を考慮撮影することができるシステムの開発を行う
- 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。

提案

- 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う
- 判定器に学習させる際、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する
- VGG16による学習を行い、3段階の評価を行う判定器を作成する



前提実験

- 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
- 柿森ら[4]が提案した料理写真の基本構図の決定方法をもとにTwitter上の料理写真の構図と「いいね」の関係性を調べた

基本構図の例





手法

- Twitter上から「定食」のキーワードで600件の画像収集を行った
- 「いいね」の数を元に、上位1～200、200～400、400～600の3つに分ける
- 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

お皿の数

順位

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

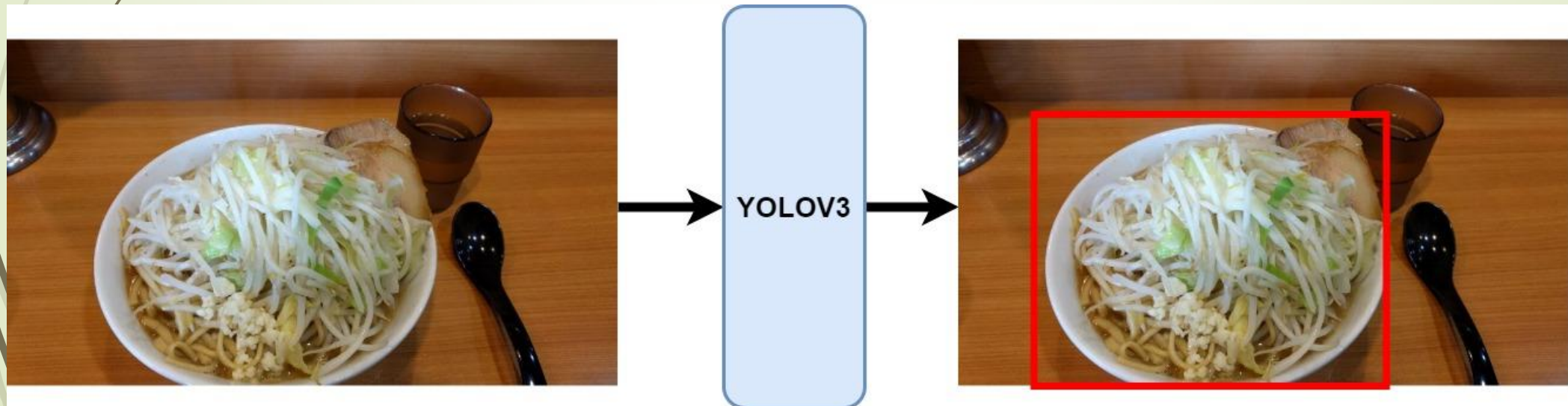


データセットの作成

- お皿が1枚の料理をデータセットとして使うため、Twitter上から「ラーメン」の画像と画像の「いいね」の数を1000件収集
- 「いいね」の数で順位ごとに並べ、順位を0から1の値に正規化
- 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
- **料理検出**を行い対象のみが写った画像に加工する

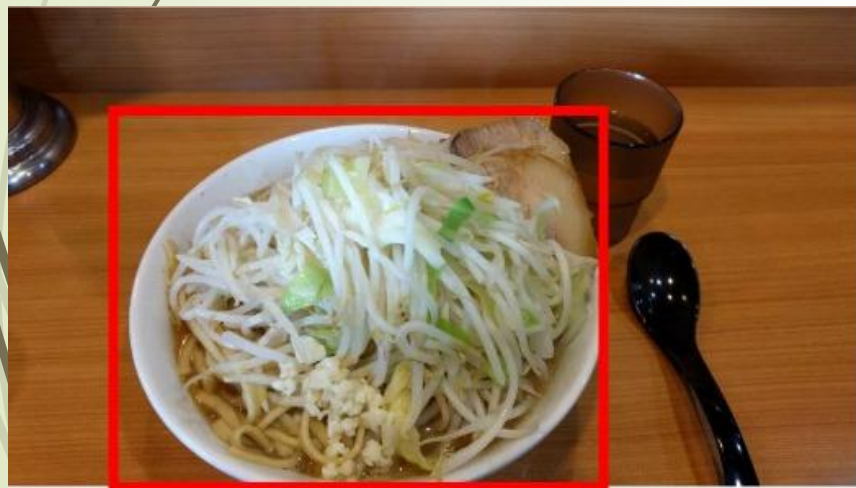
料理検出①

- 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。
- 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
- 「ラーメン」の画像100枚学習を行った



料理検出②

- 検出されたバウンディングボックスの外側を削除
- 背景が黒くなった画像をデータセットとして使用する



バウンティーボックス
の外を削除



判定器

- 判定器の作成にはVGG16に転移学習を行った
- 作成したデータセットを244×244にリサイズし入力
- データセットの画像と画像に対するラベルを入力
- 出力の全結合層のみ重みの変更を行う

ネットワークのパラメータ

- 全結合層のユニット数 : 256×2
- 全結合層の活性化関数 : ReLU ($f(u) = \max(u, 0)$)
- 出力層のユニット数 : 3
- 出力層の活性化関数 : ソフトマックス
- 最適化手法 : Adam (AdaptiveMomentEstimation)
- 評価指数 : 出力と教師信号との平均絶対誤差
(MAE : MeanAbsoluteError)
- 学習回数 : 100回

評価方法

- ① 「ラーメン」の画像を学習したYOLOV3を使いデータセットにない画像を入力を行い、料理検出が行えるか評価する。
- ② 判定器に学習を行い、トレーニングデータとテストデータのそれぞれの正解率を出力する
- ③ 判定器のみの評価を行うため料理のみの画像を入力し、出力を行う

結果①

- 「ラーメン」の画像を学習したYOLOV3を使いデータセットにない「ラーメン」の画像を入力を行う



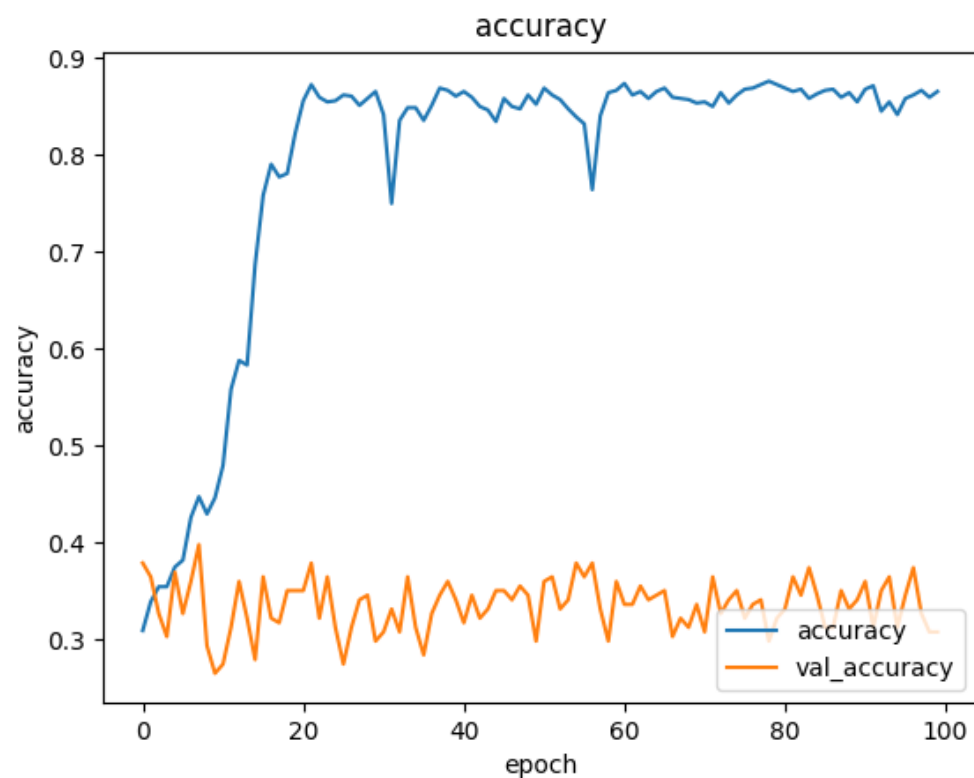
A



B

結果②

- 作成したデータセットをVGG16に転移学習を行い順位ごとに3段階に分類し、学習を行った正解率



結果③



元画像

判定器の出力：構図 = 良い



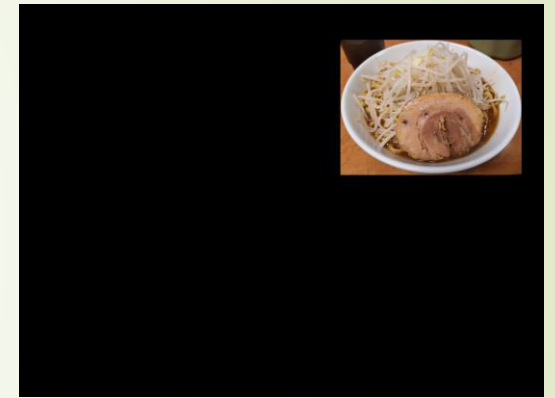
対象が中央から外れ大きく写っている画像

判定器の出力：構図 = 中程度



対象が中央で小さく写っている画像

判定器の出力：構図 = 中程度



対象が中央から外れ小さく写っている画像

判定器の出力：構図 = 悪い



考察_1

- YOLOV3を使い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
 - 対象が大きく映っている場合には高い確率で料理の検出を行うことができた。
 - 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。
-
- データセットの画像に比較的对象が大きく写った画像が多く含まれている。

考察_2

- ▶ 判定器の学習では、訓練データでの正解率は高くなった
- ▶ テストデータでの正解率を高くすることができなかった
- ▶ 実際の出力でも精度の高い判定をすることはできなかった。
- 料理検出がうまく検出が行えてない画像など学習してしまった

結論

- 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問題
 - 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
 - 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た
- Twitter上の画像と画像に対する「いいね」の値を使い判定器を作成することができた。
 - 精度の高い判定を行うことが出来なかった
- 今後データセットの改良や料理検出の精度を高めることで、実用的な判定器として使用できる。

今後の展望

- 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検出精度を高める
- 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を行う
- 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の作成
- 5つの構図の要素すべてに対応した写真撮影補助システムの作成

参考文献

- [1]角野杏早比、太田祐揮、志津野之也、原大輔、濱川礼：写真撮影における構図決定支援システム ～E-cose～
- [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
- [3]田中直人、高島健太郎：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
- [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
- [5]KazumaTakahashi、KeisukeDoman、YasutomoKawanishi、TakatsuguHirayama、IchiroIde、DaisukeDeguchi、HiroshiMurase:"EstimationoftheAttractivenessofFoodPhotographyFocusingonMainIngredients", Proc. of9thWorkshoponMultimediaforCookingandEatingActivities(CEA)、pp. 1-6、Aug. 2017
- [6]大木慧矢、永井保夫、マッキンケネスジェームス：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。情報処理学会第82回全国大会