4R-03

CycleGAN を用いた食べかけ料理画像の復元

千葉奏音　　石畑 宏明

東京工科大学コンピュータサイエンス学部

1. はじめに

本論文では画像生成モデルの一種である CycleGAN を用いて食べかけの料理の画像から食べる前の料理の画像を生成する手法について述べる．通常の GAN が，学習データと似たような新しいデータを生成するのに対して，CycleGAN では，2 種類の異なるドメイン間で画像を相互に変換することを学習する．入力として，食べる前の画像と一口食べてしまった後の画像を入力を用いて，CycleGAN を学習させ，一口食べてしまった画像から，食べる前の画像を復元することを試みる．

2. CycleGAN

GAN は，Generator と Discriminator という二つのネットワークを交互に使用することで精度の高い画像生成を行うことができる．CycleGAN は，2 組の GAN を組み合わせたものである [1].

図 1 に示すような，食べる前の画像から食べた後の画像を生成する GAN(gen_g) と図 2 に示すような，食べた後の画像から食べる前の画像を生成する GAN(gen_f) を組み合わせる．gen_g によって，入力データ bef からその食べかけ画像 aft' の生成を行う．同様に，gen_f によって，入力データ aft からその食べる前の画像 bef' の生成を行う．さらに，図 3 にあるように，bef → gen_g → aft' → gen_f → bef" , および，aft → gen_f → bef' → gen_g → aft" の変換処理を行い，bef と bef" および aft と aft" が近づくように学習させる．このようにして，双方向の変換を学習させてゆく．

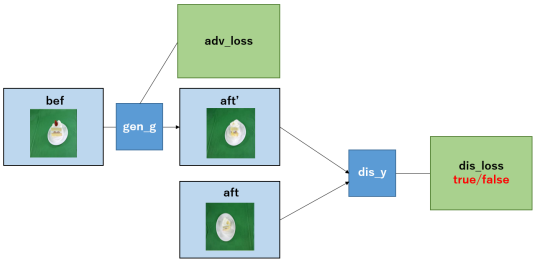

図 1: 食べる前から食べた後への変換

3. データセット

CycleGAN の学習のために，データを用意した．表 1 に使用した食べ物と画像枚数を示す．17 種類の食べ物について、複数の皿や背景を用いて画像データ撮影を行った．専用の撮影器具を用意し、Web カメラ 6 台を縦一列に並べ、グリーンバックを背景とし，テーブル上に食べ物を配置して撮影を行った．まず，食べる前の画像を撮影し，その後一口食べて，食べた後の画像の撮影を行った．なお，学習に用いた「リンゴ」のみスマートフォンで撮影した．

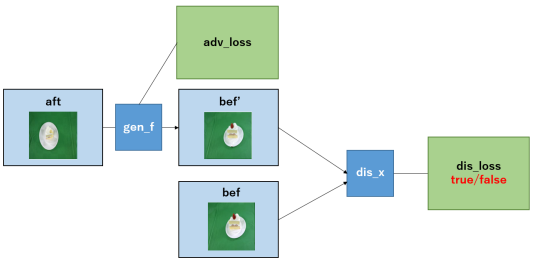

図 2: 食べた後から食べる前への変換

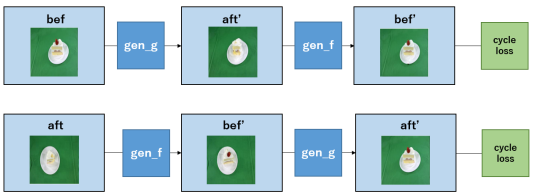

図 3: 元にもどすように学習

4. 実験

本実験では全部で 15,491 枚の画像を使用した．

iteration とは、学習を行った回数を示す．以下に学習結果を示す．図 4,5,6 について左から元画像、元画像から生成した画像、生成した画像から元画像に生成した画像を示す．

図 4 にリンゴにおける結果を示す．食べる前と後の相互に生成ができている．下段の画像について、食べる前と大きく異なっていたが食べかけ部分との境が分からないくらい食べる前の画像生成ができている．

図 5 にシュークリームと肉まんの結果を示す．上段の肉まんは食べかけ画像への生成ができなかったが下段のシュークリームは食べかけ画像から食べる前の画像を生成できた．

図 6 に二個並べたショートケーキの結果を示す．上段を見ると元画像にあった片方のケーキがなくなっていることが分かる．下段は逆に元画像でなかったもう一つのケーキのようなものが生成されている．

結果より、画像生成の精度が色に大きく影響している．また図 4 のリンゴのように外と中身が異なる色合いのものも比較的良い結果になることが分かった．一方で図 5 の上段にしようされた肉まんのように白い食べ物は白いお皿に置くと学習しづらい為うまくいかなかった．図 6 よりショートケーキの個数

Image Correction for Half-eaten dishes Using CycleGAN
Kanade CHIBA and Hiroaki ISHIHATA
School of Computer Science, Tokyo University of Technology

Copyright ©2019 Information Processing Society of Japan.
All Rights Reserved.

表 1: 撮影した食品と画像の枚数

料理名	枚数
チーズケーキ	1878
抹茶ケーキ	864
ショートケーキ	2718
クレープ	252
シュークリーム	1815
ハンバーガー	2148
たまごパン	243
あんぱん	222
肉まん	864
ピザ	456
コロッケ	438
餃子	288
焼き鳥	252
おにぎり	252
春巻	288
饅頭	180
リンゴ	2153

が変化していることから、同じ画像内に複数の食べ物を置いた場合は対象が定まらない為精度は低くなった．

図 4: リンゴ画像による結果 (iteration=119300)

5. 関連研究

成冨らは CycleGAN から派生した conditional CycleGAN を用いた食事画像変換手法の提案を行った [2]．conditional CycleGAN は、入力画像から複数種類の画像変換ができる特徴がある．計 230,053 枚の学習データを使用し、10 のカテゴリに食事画像の変換を行った．データセットあたりの学習枚数が多い方がより細かな変換ができるという結果だった．

6. まとめ

本研究では画像生成モデル CycleGAN を用いて食べかけ料理画像から食べる前の料理画像の生成を行った．主に専用の撮影器具を用いて 17 種類の食べ物についてデータセットを作成した．結果から、背景と色のはっきりと異なる食べ物や、食べた後で外と中身の色が異なる食べ物については、比較的自然な

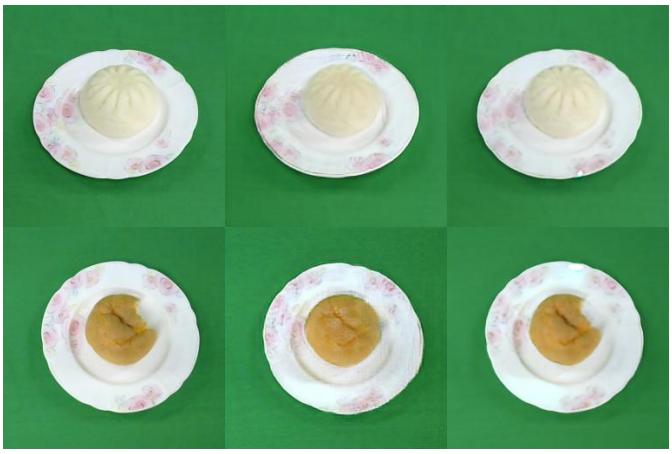

図 5: 肉まん（上段）とシュークリーム（下段）による結果 (iteration=44000)

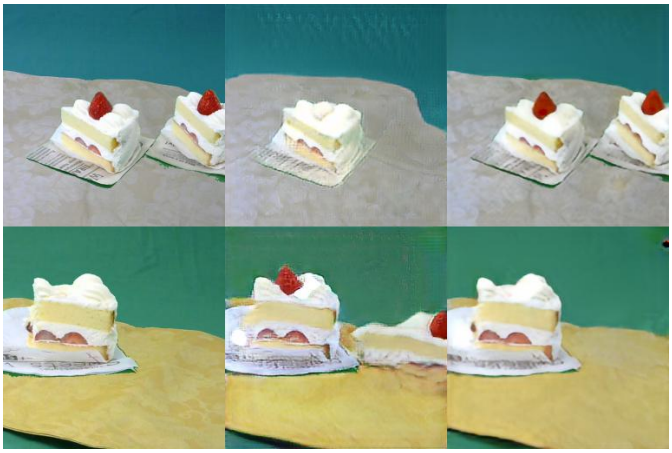

図 6: 2 つのショートケーキによる結果 (iteration=49000)

画像が生成できた．複数の食べ物を用いて画像の生成を行う場合は、ネットワークの改善が必要である．

参考文献

[1] Jun-YanZhu,Taesung Park,Phillip Isola,Alexei A.Efros “Unpaired Image-to Translation using Cycle-Consistent Adversarial Networks”，ICCV 2017paper，pp.2223-2232，Feb.19，2018

[2] 成冨志優, 堀田大地, 丹野良介, 下田和, 柳井啓司 “Conditional GAN による食事写真の属性操作”, DEIM Forum 2018 F1-3

Copyright ©2019 Information Processing Society of Japan.
All Rights Reserved.