

Conditional GAN を用いた複数詳細カテゴリ画像の合成

伊藤 祥文^{1,a)} Jaehyeong Cho^{1,b)} 柳井 啓司^{1,c)}

1. はじめに

現在、深層学習技術の発展により、画像認識の精度が向上したことで、画像を用いた様々な研究が行われている。特に、画像認識分野においては、Deep Learning を用いて人間の能力以上の結果を出すことに成功しているが、それには大量の画像が必要になるため、詳細画像認識は現在も研究が進められている分野となっている。また、2014年に発表された GAN(Generative Adversarial Network)[1] という手法を用いて自動画像生成が行われたことが契機となり、近年、自動画像生成がさかに行われるようになった。

よって、本研究では、GAN によって自動で生成された画像を評価し、それを用いてデータセットを作成することを考える。これにより、これまで詳細画像認識を行う上での障害となっていた、学習用画像の不足を補うことができると考えられる。

2. 目的

本研究では、GAN の一種である Conditional GAN[2] を用いて画像生成を行い、生成された画像を評価するとともに、それを用いてデータセットを構築することを目的とする。また、既存のカテゴリに属する画像を用いて新しいカテゴリを作り、画像データセットを拡張する。本研究では、食事画像、特にラーメンの画像を扱い、特定のラーメンの画像を自動で生成した。

3. 関連研究

3.1 画像認識および分類

近年の画像認識に関する研究の一つとして、大量の画像を分類する研究が行われている。例えば、Yang らの研究 [3] では、車の画像のデータセットを作成し、それを用いて一般の車の画像を分類する手法を提案している。

車以外の画像を分類した例として、松田らの研究がある [4]。この研究では、大量の食事画像を用いて UEC-FOOD100 食事画像データセットを作成している。

本研究では、この研究で作成されたデータセットを用いてラーメン画像の収集を行い、画像生成を行う。

3.2 画像の自動生成

近年、GAN を用いた画像の自動生成に関する研究は盛んに行われている。例えば、Radford らの研究 [5] では、GAN を派生させた手法である DCGAN を構築し、より本物に近い画像の生成に成功している。この研究では、CNN (Convolutional Neural Network) を用いることで画像認識の精度を向上させるものである。また、Zhang らの研究 [6] では、GAN を応用して説明文から画像を生成させる StackGAN を作成している。この研究では、GAN を複数重ねることにより、高品質な画像の生成に成功している。

本研究では、Conditional GAN[2] を用いて画像生成を行う。この手法は、GAN の generator と discriminator にカテゴリラベルを設定することにより、以下の式を用いてより高品質な画像を生成するものである。

$$\min_G \max_D V(D, G) = E_{x \sim p_{data}(x)} [\log D(x|y)] + E_{z \sim p_z(z)} [\log(1 - D(G(z|y)))] \quad (1)$$

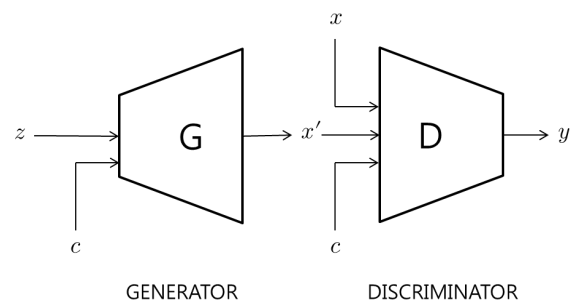


図 1 Conditional GAN のネットワーク図

Conditional GAN のネットワークは、図 1 のようになっている。z にカテゴリの condition をランダムに与えることで画像を生成し、図 2 の generator を用いて画像を生成、図 3 の discriminator により生成された画像が実画像との比較を行うものとなっている。

¹ 電気通信大学大学院 情報理工学研究所

a) ito-y@mm.inf.uec.ac.jp

b) cho@mm.inf.uec.ac.jp

c) yanai@mm.inf.uec.ac.jp

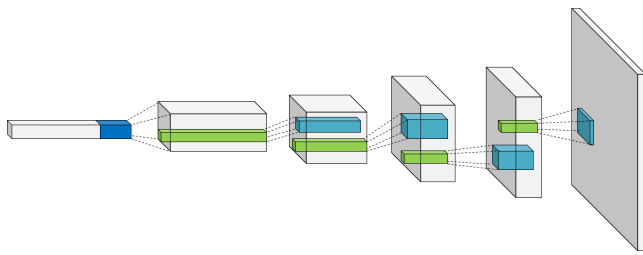


図 2 generator のネットワーク図

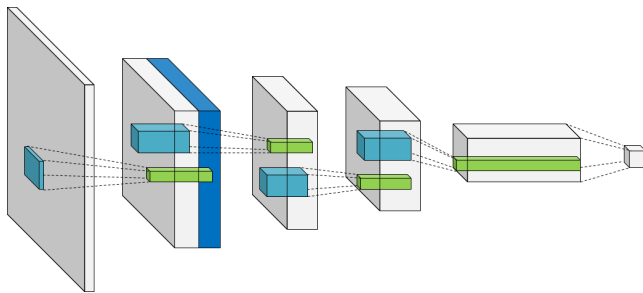


図 3 discriminator のネットワーク図

4. 手法

本研究の手法の全体的な流れは、以下のようになっている。

- (1) 画像の収集
- (2) 収集された画像を用いたカテゴリ画像の自動生成
- (3) 複数カテゴリの画像を用いた合成画像の生成
- (4) 生成された画像の評価

4.1 画像の収集

2011 年 5 月から 2016 年 4 月までの 5 年分の Twitter の過去ログから、特定の単語を含む食事画像付きツイートを収集する。得られた画像のうち、特に食事画像であるとされる画像を、食事画像データセットを用いて選別し、収集する。

4.2 画像の生成

収集された画像を用いて画像の生成を行う。本研究では、複数カテゴリの画像を用いて画像生成を行うため、Conditional GAN[2] を用いて画像生成を行う。

4.3 画像の合成

画像生成により得られたデータを用いて、複数カテゴリを用いた合成画像の生成を行う。本研究では、図 1 の z の値を固定し、その値を 0.1 刻みで変化させていくことで、あるカテゴリから別のカテゴリへの画像の連続変化を見る。また、図 1 の c の値を固定し、2 種類のカテゴリの情報のみを用いて画像生成を行う。

4.4 画像の評価

生成された画像を表示し、どのような画像が生成されたのかを目視で確認する。また、2 つのカテゴリの画像の連続変化を見ることで、生成された合成画像が実画像に近いものであるかを見る。なお、本研究で表示される画像のサイズは、 64×64 である。

5. 実験

本研究では、2011 年 5 月から 2016 年 4 月までの 5 年分の Twitter の過去ログを用いて実験を行った。

実験では、ラーメンという単語で検索を行った後、表 1 にある単語および関連する単語を含む投稿から、ラーメンの画像を選別して収集した。

表 1 対象としたカテゴリの詳細

対象としたカテゴリ	画像枚数
ラーメン二郎	5772 枚
家系ラーメン	3703 枚
台湾ラーメン	2166 枚
北極ラーメン	4482 枚
尾道ラーメン	1548 枚
その他	1010 枚
合計	18681 枚

得られた画像を用いてカテゴリごとに画像生成を行った結果、図 4 のような画像が得られた。



図 4 生成された画像の例 (上段: ラーメン二郎 中段: 台湾ラーメン 下段: 北極ラーメン)

また、2 つのカテゴリ間で画像の連続変化を行った結果、図 5 のような結果が得られた。このうち、特に両方のカテゴリの特徴を持つと考えられる画像を集めたものが図 6 である。

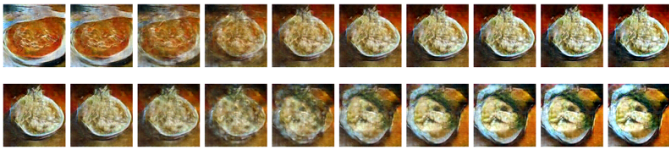


図 5 2 カテゴリ間における画像の連続変化の例
(上段左: 北極ラーメン 上段右: ラーメン二郎
下段左: ラーメン二郎 下段右: 家系ラーメン)

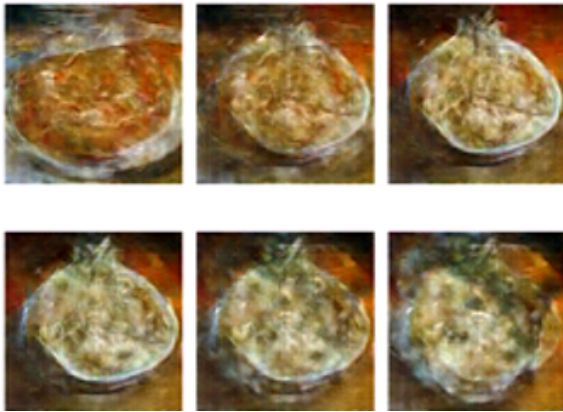


図 6 図 5 のうち、特に両方のカテゴリの特徴が見られると考えられる画像 (上: 北極ラーメン + ラーメン二郎 下: ラーメン二郎 + 家系ラーメン)

また、condition を固定し、2つのカテゴリのみを用いて画像生成を行った結果、図 7 のような画像が生成された。

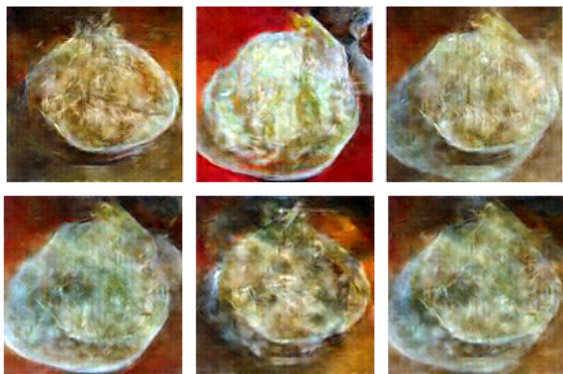


図 7 condition を固定して生成した画像 (上: 北極ラーメン + ラーメン二郎 下: ラーメン二郎 + 家系ラーメン)

6. 考察

Twitter の投稿内容という、ノイズがあるようなものをデータセットとして用いて画像生成を行ったが、ある程度実画像に近い画像データを生成することができた。しかし、実画像に似ていない画像や同じような画像、不自然な画像が生成されるなど、画像生成に失敗した例も見られた。これは、データセット内に同じ画像が含まれていたことによる学習ミスや、画像データの数が足りないことによる過学習、画像ごとのどんぶりの形状が異なっていることによる器の形の学習ミスが原因であると考えられる。

また、2つのカテゴリを用いた画像合成による新しい画像の生成の結果、図 6 のような、2 カテゴリ両方の特徴を持った新画像の生成ができた。同様に、condition を固定し、 z のみをランダムにして画像生成を行った場合も、図 7 のように、2 カテゴリの特徴を持っている画像が生成された。今後、カテゴリ数を増やした場合でも、合成画像を生成することができると考えられる。

今後は、より多くのカテゴリの画像を用いて、様々な画像の生成を目指すとともに、より実物に近い画像を生成していきたい。特に、生成される画像のサイズが 64×64 のため、小さい画像しか生成されないという問題がある。これは、StackGAN[6] の理論を用いて生成画像の解像度を上げるなど、他の GAN を応用した論文を参考にすることで解決していきたい。また、生成画像のどんぶりの形状が不自然になってしまう点が大きな問題となっている。これについては、画像内の同じ場所にラーメンのどんぶりがあるような画像をあらかじめ選別するなど、どんぶりの形状に関する制約を導入した Ramen GAN を検討中である。

7. 終わりに

我々はより詳細な食事画像データセットの作成を目標としている。今回の結果から、Conditional GAN の condition や変数 z の与え方を変えることで、未知のカテゴリについての画像を生成することが可能であることがわかった。今後はデータ量の少ないカテゴリや、全く新しいカテゴリについても画像生成を行い、データの補強を行えるようにしていきたい。

参考文献

- [1] Goodfellow, Ian and Pouget-Abadie, Jean and Mirza, Mehdi and Xu, Bing and Warde-Farley, David and Ozair, Sherjil and Courville, Aaron and Bengio, Yoshua, Generative adversarial nets Advances in neural information processing systems, pp. 2672–2680, 2014.
- [2] Mirza, Mehdi and Osindero, Simon, Conditional generative adversarial nets, arXiv preprint arXiv:1411.1784, 2014.
- [3] Yang, L. and Luo, P. and Loy, C. C. and Tang, X., A Large-Scale Car Dataset for Fine-Grained Categorization and Verification, Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 3973–3981, 2015.
- [4] Matsuda, Y. and Hoashi, H. and Yanai, K., Recognition of Multiple-Food Images by Detecting Candidate Regions, Proc. of IEEE International Conference on Multimedia and Expo (ICME), pp. 25–30, 2012.
- [5] Radford, A. and Metz, L. and Chintala, S., Unsupervised Representation Learning with Deep Convolutional Generative Adversarial Networks, Proc. of ICLR, 2016.
- [6] Zhang, H. and Xu, T. and Li, H. and Zhang, S. and Huang, X. and Wang, X. and Metaxas, D., StackGAN: Text to Photo-realistic Image Synthesis with Stacked Generative Adversarial Networks, Proc. of arXiv:1612.03242, 2016.