

# GANによる画像生成を用いたデータ拡張に基づく物体検出

坂本 悠哉<sup>†</sup> 大川 茂樹<sup>†</sup>

千葉工業大学 先進工学研究科 未来ロボティクス専攻<sup>‡</sup>

## 1. はじめに

昨今、機械学習やディープラーニングによる画像認識技術が発展しており、リアルタイムでも高い精度での物体検出が可能となっている。またディープラーニングによる物体検出では、データ数が多くなると精度が上がる傾向にあるが、高い精度を得るには数千数万のデータが必要となる。これに対処する手段として、データ拡張がある。データ拡張の一般的な手法として、画像の回転、反転、拡大、縮小がある。本研究では、データ拡張の一手法として、GAN(Generative Adversarial Network)を用いた画像生成を採用する[1]。GANを使用したデータ拡張では、高画質で綺麗に生成された画像を用いる研究は多く存在する。一方で、形状が歪んで生成された画像をデータ拡張に加える研究は少ない。そこで本研究では、生成された画像の中でも形状が歪んでいるものに焦点を当て、これをもとにデータ拡張をおこなった。

## 2. 目的

形状が歪んで生成された画像を利用したデータ拡張による物体検出をおこない、本研究で提案するデータ拡張手法の有用性について調査する。

## 3. 方法

### 3.1 データセット

本研究では、The Oxford-IIIT Pet Dataset[1]をデータセットとして使用した。The Oxford-IIIT Pet Datasetは犬と猫の37クラスの血統が1クラスにつき200枚含まれている。この内、アメリカンブルドックとパグの画像を使用した。

### 3.2 DCGANによる画像生成

画像生成にはDCGAN(Deep Convolutional Generative Adversarial Network)を用いる。この手法はCNNで実装された生成機と判別機を共進化させることで高性能の生成機を得ている[1]。

また画像生成用のデータセットとして3.1で挙げたアメリカンブルドックとパグの画像から顔領域を切り出す前処理をおこなった。

### 3.3 SwinIRによる画像の高画質化

3.2にて顔領域を切り出した画像にSwinIR[2]により画像の高画質化をおこなう。SwinIRは2021年に発表されたSwin Transformerに基づくモデルであり、浅い特徴抽出、深い特徴抽出、高画質の画像再構成の3つの部分で構成されている。

### 3.4 YOLOv5による学習

物体検出の方法として、近年物体検出の主な手法として挙げられるYOLO(You Only Look Once)v5[3]を使用する。YOLOシリーズの初代であるYOLOv1をもとに改良が続き、YOLOv5では、非常に速く、正確な検出が可能である。

### 3.5 Grad-CAMによる特徴分析

Grad-CAM(Gradient-weighted Class Activation Mapping)[4]は、入力画像に対してCNNモデルがどの領域に着目しているかを可視化出来る手法である。この手法を用い生成画像の特徴分析をおこなった。

## 4. 結果

### 4.1 DCGANによる画像生成結果

アメリカンブルドックとパグの画像を200枚ずつ用意し、3.2、3.3で触れた前処理をおこなったデータセットを用いDCGANによる画像を生成する。Fig. 1は形状が歪んで生成された画像例である。



Fig. 1 DCGANによる画像生成例

## 4.2 Grad-CAMによる特徴分析

4.1 で集めた形状が歪んで生成された画像を Grad-CAM を用いて特徴分析をおこなった. 特徴分析をおこなった例が Fig. 2 である. この内犬であると判断された部分が強調されたものを物体検出のデータ用として集めた.

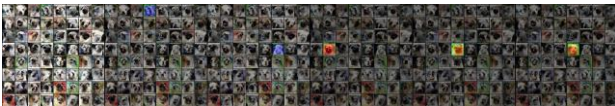


Fig. 2 Grad-CAM での特徴分析の例

## 4.3 YOLOv5 による物体検出

Table 1 で示すように, データ数は 3.1 で集めた画像 280 枚を Train データ, 80 枚を Val データとして学習をおこなった. また, Train データ 280 枚の内 4.2 で集めた画像を含んだデータでも学習をおこなった. Test データは 40 枚とした.

学習した重みを用いて Test データ内の画像から犬の検出をおこなった. ここで画像内の全ての犬を検出が可能となった場合を Positive とし, 犬を検出出来ていない場合は Negative とした. Test データの Positive の割合を Accuracy とした. 検出時の confidence は初期設定である 0.5 以上のバウンディングボックスを表示した.

Table 2 は YOLOv5 で学習した重みで検出をおこなった結果であり, 形状が歪んで生成された画像を含んだデータで学習した場合でも Accuracy は 0.90 となった.

Table 1 学習に使用するデータ

Train (枚)	Val (枚)	Test (枚)
280	80	40
280*	80	40
280**	80	40

\* 280 枚の内 70 枚は GAN による生成画像

\*\* 280 枚の内 140 枚は GAN による生成画像

Table 2 YOLOv5 による検出結果

Train (枚)	Val (枚)	Test (枚)	Accuracy
280	80	40	0.95
280*	80	40	0.85
280**	80	40	0.90

\* 280 枚の内 70 枚は GAN による生成画像

\*\* 280 枚の内 140 枚は GAN による生成画像

## 5. おわりに

本研究では, DCGAN による画像生成により, 形状が歪んで生成された画像でデータを拡張し, 物体検出をおこなった. その結果, Train データの半分以上が, 形状が歪んで生成された画像であっても Accuracy は 0.90 という結果が得られた. 今後はデータ数が多い場合での検証をおこない多数のデータでも高い精度の維持を目指す.

## 参考文献

- [1] Hiroki Watabe, Hiroshi Watanabe, "DATA AUGMENTATION USING DCGAN FOR BREED IDENTIFICATION", Proceedings of the Firth IIEEJ International Workshop on Image Electronics and Visual Computing 2017
- [2] Jingyun Liang, Jiezhong Cao, Guolei Sun, Kai Zhang, Luc Van Gool, Radu Timofte, "SwinIR: Image Restoration Using Swin Transformer", 2021 IEEE/CVF International Conference on Computer Vision Workshops (ICCVW)
- [3] Liu Yifan, Lu BingHang, Peng Jingyu, Zhang Zihao, "Research on the Use of YOLOv5 Object Detection Algorithm in Mask Wearing Recognition", World Scientific Research Journal, 276-284, (2020)
- [4] Ramprasaath R. Selvaraju, Michael Cogswell, Abhishek Das, Ramakrishna Vedantam, Devi Parikh, Dhruv Batra, "Grad-CAM: Visual Explanations from Deep Networks via Gradient-based Localization", International Journal of Computer Vision (IJCV) in 2019