

GANomaly と動物の画像を使った異常行動の検出

高久優典[†] 田村仁[†]日本工業大学先進工学部 ロボティクス学科[†]

1. はじめに

動物園や野生動物の監視において、カメラ映像から状態を自動で判別することが作業を効率化するうえで必要とされている。状態を判別する方法として、機械学習による画像分類がある。画像分類とは、画像の中にあるモノをカテゴリ分けする技術である。

特に、動物の監視においては、異常行動を検出したい。例えば、歩いている姿や座っている姿のような日常的に見る行動を正常行動とし、嘔吐や失神をするような非日常的な行動を、ここでは異常行動と呼ぶことにする。異常行動は一般的には稀な行動であるため、正常な行動と異常な行動をカテゴリ分けすると、データ数に偏りが発生してしまう。しかし、機械学習による画像分類では、各カテゴリのデータ数をおおむね均等にする必要はある。

このようなデータに偏りがある場合に適した方法として異常検知という技術がある。異常検知は、データの中から特徴的な外れ値を検出する。つまり、多数を占める正常行動の画像だけを学習させ、その平均的な画像から大きく外れる画像を異常行動として検出する。

これを行うには、多数の画像から平均的な画像を生成する必要がある。その一方で、最近の機械学習分野では敵対的生成ネットワーク (Generative Adversarial Networks 略称:GAN) [1] による画像合成手法が盛んに研究されている。この GAN を平均的な正常画像の合成に用いた手法として AnoGAN[2]がある。AnoGAN は部品の傷などの検出を目的としており、平均的な正常画像を 1 種類だけ生成する。そのため、入力画像に動物の行動のようなバリエーションが多いものを使用すると、対応できない。それに対して GANomaly[3]では、入力画像に対応する正常な合成画像を、複数種類生成することができる。本研究では、この GANomaly を用いて動物の異常状態の検出ができるかを研究した。

2. GANomaly

GANomaly は、画像生成に DCGAN[4]を用い

ている。DCGAN[4]は Pix2Pix[5]や CycleGAN[6]のような画風変換に用いられる。DCGAN は入力画像に対応した再構築画像を生成する。GANomaly は、学習した結果、正常画像に近似した再構築画像を生成する。異常画像の場合は、異常な部分を再現できずに正常画像に寄せた歪な再構築画像を生成する。入力画像と再構築画像との差は異常な画像ほど大きくなるため、その差を元に判定する。

3. 実験方法

本研究では、対象を犬とした。犬の動画や画像はインターネットに多く存在しており収集が容易である。動画の場合は、フレームごとに分けた。本研究では、歩くや座るなどの日常的にする行動を正常、嘔吐する行動を異常とする。横から見て犬の全身が映っているような画像を集めた。これは真上などから見た画像では犬かどうかは人間でも判別することすら難しいためである。また学習するにあたって、画像から犬のみを切り抜いて 2 値化した。これは背景の影響と、体の色によって学習結果に影響が出ないようにするためである。画像サイズは大量に学習する上で時間をかけ過ぎず、人間が見ても犬の姿勢は判断できるように 128x128 (pix) にリサイズした。結果的に用意できたデータは、訓練用画像は 13,888 枚、テスト用画像は正常が 3,999 枚、異常が 1,350 枚となった。元のプログラムはテスト用画像を正常異常のどちらかに判別したか分からないため、一部を書き換えて正常とした画像を正常ファイル、異常とした画像を異常ファイルに分けられるようにした。

4. 結果

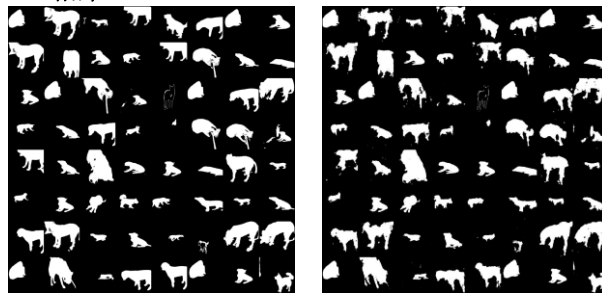


図1 (a) は入力画像の一部 (8x8 の 64 枚)、(b) は(a) を元に生成されたフェイク画像 (同じく 64 枚)

Pose Estimation with images of animals and GANomaly

[†]Yusuke Takaku, Hitoshi Tamura

Department of Robotics, Faculty of Advanced Engineering,
Nippon Institute of Technology

図1の画像は8x8の64枚の画像で生成されているが、これはプログラム上のバッチサイズに依存しており、本研究ではバッチサイズは64である。元画像と再構築画像を比較すると、元画像に似た犬画像を生成して騙そうとしているのが分かる。

表1 正常と異常の割合（目視による概算）

		予測値	
		P	N
真値	P	3,580	419
	N	62	1,288

表1は、目視による概算での混同行列である。Pは正常、Nは異常の枚数を示す。表1によると、異常と判定された正常画像が419枚で、全正常画像のうち約11%、正常と判定された異常画像は62枚で、全異常画像のうち約5%であった。異常と判定された正常画像の多数は立っている状態で、歩いている状態や伏せている状態はわずかであった。

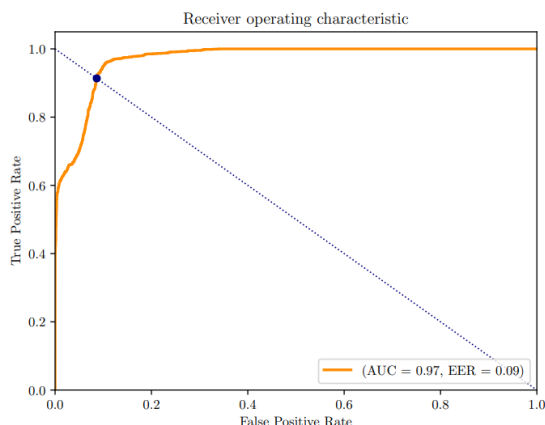


図2 GANomalyのROC曲線

ROC曲線とは、分類問題における評価指標の1つである。全てのNegativeのうち、誤ってPositiveと判定した割合であるFPR(False Positive Rate)を横軸、全てのPositiveのうち、正しくPositiveと判定した割合であるTPR(True Positive Rate)を縦軸としており、PositiveとNegativeを分類する閾値を変更していった際の、TPRとFPRの値を取っている。学習結果の精度が良いほど、曲線は左上に歪曲していく。AUCとは、このROC曲線の下側の面積のことであり、この面積が大きいほど分類がうまくいっている。図2によると、ROC曲線は大きく左上に膨らんでおり、AUCも0.97と精度は高いといえる結果となった。

4. 考察

動物は嘔吐する前に、嗚咽を数回ほど繰り返

す。人間はこの動作を確認した時、この後に嘔吐することを予測することができる。しかし、GANomalyは嗚咽の動作を1フレームの画像として見ているため、動物が立ち止まっている状態と判定して正常扱いになったと考える。そのため、1フレームの前後の情報を含めて学習すれば、うまく判定できるようになる可能性がある。歩行や伏せている状態はうまく判定できていたが、誤って判断された画像は、立ち止まっている状態や座っている状態の画像が多かった。これは、立ち止まっている状態、座っている状態は正常にも異常にも近い状態のため誤認した可能性がある。本研究では対象にしていなかった失神のような、意識を失って倒れている状態を判定させたとき、伏せていると判定されると予想される。学習に用いた画像の中には2値化に失敗している画像もあったため、それらが悪影響を及ぼしていると考えられる。

5. おわりに

本研究では、GANomalyを用いて動物の異常行動の検出を行った。その結果、立ち止まっている画像の多くが誤って判定されたが、歩行状態や、座っている状態はうまく判定できていた。本研究のような画像認識ではなく、1フレームの前後の情報も含めて学習する動画認識を使えば、精度を上げることができる可能性がある。

参考文献

- [1]Tim Salimans, Ian Goodfellow, Wojciech Zaremba, Vicki Cheung, Alec Radford and Xi Chen, "Improved Techniques for Training GANs", Advances in Neural Information Processing Systems. pp. 2234-2242, 2016
- [2]Thomas Schlegl, Philipp Seeböck, Sebastian M.Waldstein, Ursula Schmidt-Erfurth and Georg Langs, "Unsupervised Anomaly Detection with Generative Adversarial Networks to Guide Marker Discovery", Lecture Notes in Computer Science(LNCS). pp. 146-147, 2017
- [3]Samet Akcay, Amir Atapour-Abarghouei and Toby P. Breckon, "GANomaly: Semi-Supervised Anomaly Detection via Adversarial Training", [arXiv:1805.06725v3](https://arxiv.org/abs/1805.06725v3) [cs.CV], 2018
- [4]Alec Radford, Luke Metz and Soumith Chintala, "Unsupervised Representation Learning with Deep Convolutional Generative Adversarial Networks", [arXiv:1511.06434](https://arxiv.org/abs/1511.06434)[cs.LG], 2016
- [5]Phillip Isola, Jun-Yan Zhu, Tinghui Zhou and Alexei A. Efros, "Image-to-Image Translation with Conditional Adversarial Networks", IEEE Conference on Computer Vision and Pattern Recognition(CVPR). pp. 5967-5976, 2018
- [6]Jun-Yan Zhu, Taesung Park, Phillip Isola and Alexei A. Efros, "Unpaired Image-to-Image Translation using Cycle-Consistent Adversarial Networks", IEEE Conference on computer vision. pp. 2223-2232, 2017