

# 実画像データが少ないなかでのAI開発 に対する考察

～sim2realの動向調査(2021年10月時点)～

ICTS開セ／FA開セ／デジタルソリューション部・新入社員  
尾崎 亮太(Ozaki, Ryota)

## 1. 緒言

近年、建設機械／現場でDeep learningを活用する研究開発が進められている。一般的に、DNN(deep neural network)を訓練するためには、大量の訓練データが必要になる。例えば、建設機械周辺の人検知をDNNで行うためには、大量の車載カメラ画像が必要になるだろう。しかしながら、建設現場のカメラ画像を大量に収集するのは容易ではない。なぜなら、お客様の機械／現場にカメラを設置するための交渉や、データをサーバにアップロードするための通信環境の整備が必要になるからだ。そのため、現状では、限られたRealデータでDNNを訓練するための工夫が必要になる。

上記のことをふまえて、本レポートでは、少量のReal画像データでDNNを訓練するためのアプローチについて文献調査を行い、考察をまとめることとする。ただし、本調査は2週間程度で行われており、偏った調査になってしまっている可能性があることに留意すべきである。

## 2. 調査結果

少量のReal画像データでDNNを訓練するためのアプローチについての文献調査結果を以下にまとめる。

### 2-1. 少量のRealデータでの学習テクニック

少ないRealデータでの学習テクニックを以下に挙げる。

- データ水増し(data augmentation)：  
データを加工する(e.g. 画像の反転、回転)ことで訓練データの量を増やす。
- 転移学習(transfer learning)：  
他のタスクで訓練されたネットワークを転用する。
- Fine-tuning：  
Realデータでネットワークを訓練する前に、Simデータで事前学習を行う。
- sim2real(sim-to-real)：  
Simで得られた知識をRealに適用する(ときのギャップを埋める)。
  - 環境乱択性(domain randomization)：  
学習するSim環境をランダム化することで、sim2realのギャップを埋める。
  - ドメイン適応(domain adaptation)：  
ドメインが異なる2つのデータセット(SimとReal)を同じように扱うために変換などを行う。

上記のうち、本レポートでは、ドメイン適応に着目する。なぜなら、容易に大量収集できるSimデータを、Realそっくりなデータに変換できれば、訓練データ不足を根本的に解決できると考えるからだ。

### 2-2. ドメイン適応の動向

ドメイン適応の動向を以下でまとめる。

#### 2-2-1. GAN(2014)

GAN(generative adversarial network)<sup>[1]</sup>は、用意されたデータ群からその特徴を学習することで、擬似的なデータを生成することができる。図1にその例を示す。GANは、生成器(generator)と識別器(discriminator)で構成

される。生成器は、本物のような疑似画像を生成できるようになることを目的とする。一方で識別器は、本物と疑似画像を区別できるようになることを目的とする。これらを競合させて、交互に訓練することで、生成器は本物そっくりの疑似画像を生成できるようになる。

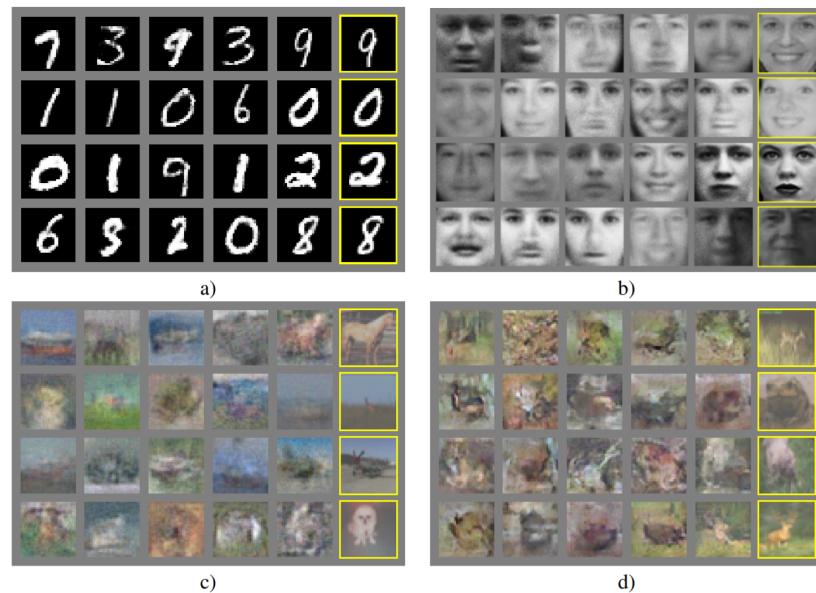


図1:訓練データの例(右端列、黄色枠)とGANによって生成された画像の例  
(文献[1]より転載)

(初期の)GANの特徴を以下にまとめると。

- 訓練データと似たような画像を生成できる。
- ✗ ドメインの変換はできない。

## 2-2-2. Pix2Pix(2017)

Pix2Pix<sup>[2]</sup>は、異なる2つのドメインの画像をペアとして学習することで、入力画像を別の性質(ドメイン)に変換することができる。図2にその例を示す。この手法では、変換前後の画像ペアを大量に用意し、GANを用いてその変換を再現できるように学習する。

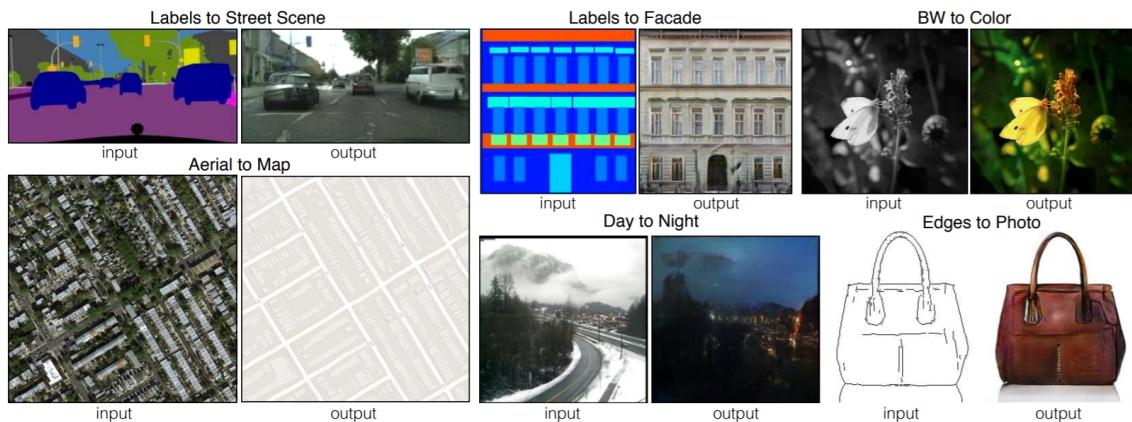


図2:Pix2Pixの入出力例  
(文献[2]より転載)

Pix2Pixの特徴を以下にまとめると。

- 訓練データを画像ペアにして学習することで、入力画像を別の性質へ変換することができるようになる。
- ✗ 訓練データとして、輪郭が一致している1対1の画像ペアを大量に用意する必要があり、それは容易ではない。

## 2-2-3. CycleGAN(2017)

CycleGAN<sup>[3]</sup>は、Pix2Pixと同様に、入力画像を別の性質(ドメイン)に変換することができる。図3にその例を示す。

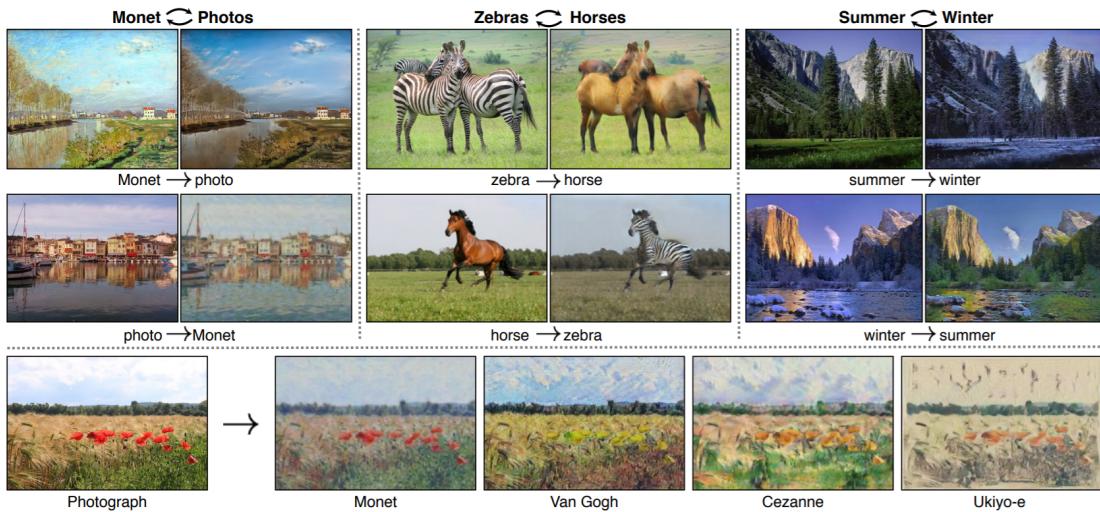


図3:CycleGANの入出力例

(文献[3]より転載)

Pix2Pixとの差異は、用意する訓練データセットの要求にある。前項のとおり、Pix2Pixでは、輪郭が一致している1対1の画像ペアが必要である。一方でCycleGANでは、画像がペアになっている必要はない。それぞれのドメインで大量の画像があれば、画像内の物体の形状や位置がバラバラでも学習できる。図4の例では、馬データセットXとシマウマデータセットYを用意しているが、それぞれの動物の形状、向き、数などはペアになっていない。

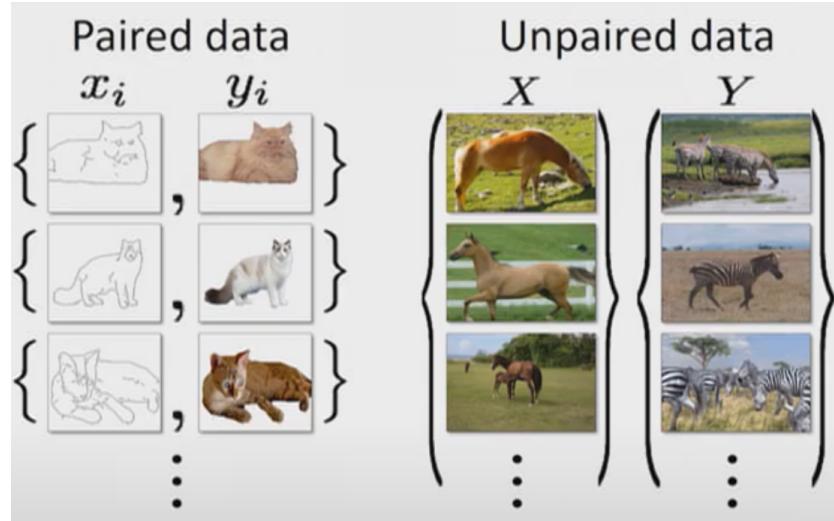


図4: Pix2Pixの訓練データ例(左)とCycleGANの訓練データ例(右)

(文献[3]より転載)

そのようなデータセットでの学習をしているのは、それぞれのドメインへの写像を”サイクル”させる仕組みである。CycleGANでは、図5(a)のように、ドメインX→Yの写像を行う生成器G、ドメインY→Xの写像を行う生成器F、各ドメインの識別器D\_X, D\_Yで構成される。そして、図5(b)のように、元画像xと、それに写像X→Y→Xを適用した生成画像x^が同一になるように、cycle consistency lossという損失関数を採用する。従来のGANの損失関数に、このcycle consistency lossを加えることで、ペアでないデータセットでも、生成の制約をつけながら学習を進めることができる。

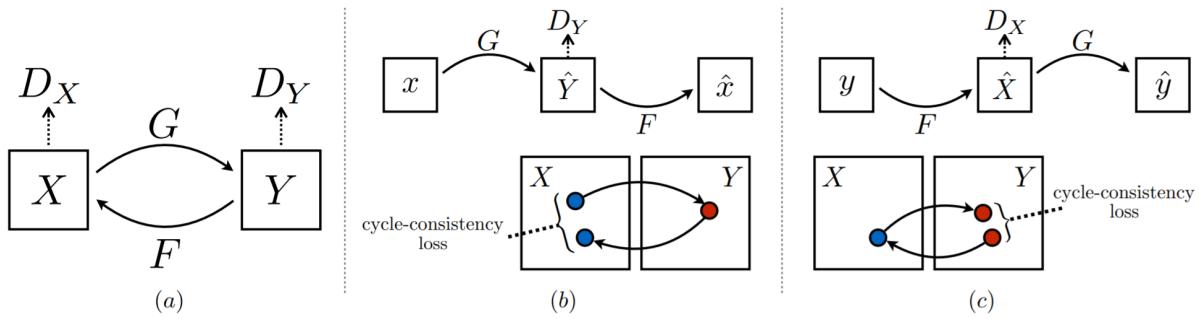


図5: CycleGANの概要  
(文献[3]より転載)

CycleGANの特徴を以下にまとめる。

- 訓練データがペアである必要がない。
- ✗ タスク(e.g. 物体の把持)に必要な詳細な情報が、変換によって失われる(e.g. 画像内の物体が歪んだり位置がズレたりする)可能性がある。

## 2-2-4. RL-CycleGAN(2020)

RL-CycleGAN<sup>[4]</sup>では、CycleGANをRL(強化学習)モデルと連結して訓練することで、CycleGANが、タスクにとって重要な情報を保持するような変換ができるようになる。この手法では、図5のように、Q値が画像変換に対して不变となるようにするRL-scene consistency lossを導入している。例えば、Q値をシミュレータ画像xから推定するQ\_simと実画像yから推定するQ\_realを用意、sim2real変換をG、real2sim変換をFとし、Q\_sim(x), Q\_real(G(x)), Q\_sim(F(G(x)))の各Q値が類似するような損失を構成する。

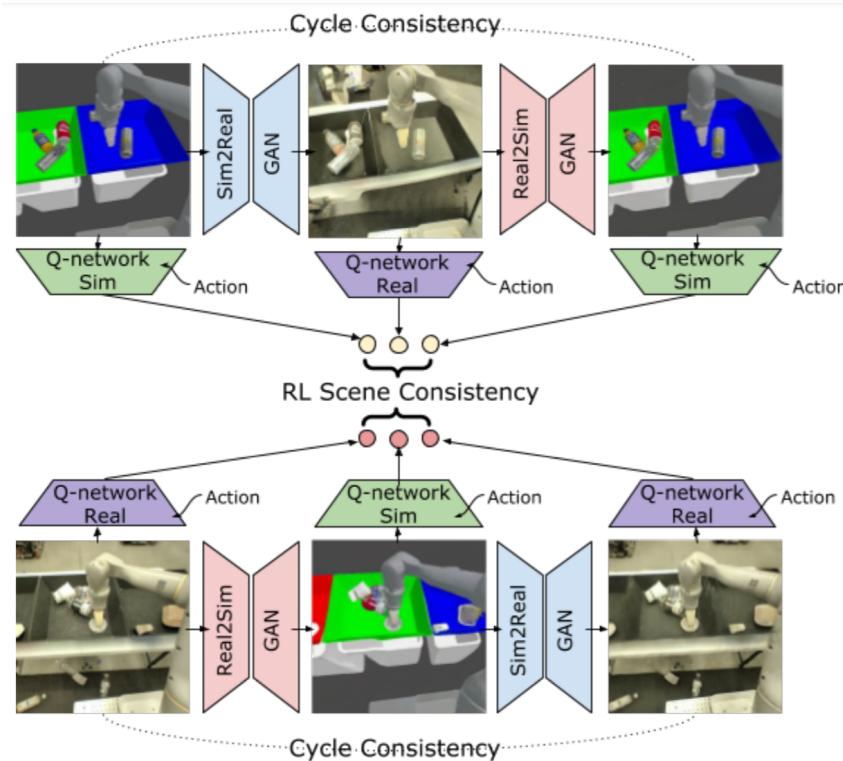


図5: RL-CycleGANの概要  
(文献[4]より転載)

RL-CycleGANの特徴を以下にまとめる。

- 特定のタスクに有用な情報を保持するようなsim2real変換ができ、タスクごとの人手によるエンジニアリングの必要性を除外できる。
- ✗ 特定のタスクに個有してしまうため、他のタスクに転用するためには学習しなおす必要がある。

## 2-2-5. RetinaGAN(2021)

RetinaGAN<sup>[5]</sup>では、sim2real変換前後の画像に対する物体検出の結果が一致するようにCycleGANを訓練することで、物体情報を把握(object-aware)したような変換ができるようになる。つまり、画像のsim2real変換前後で、画像内の物体の構造や位置が変わらないようにできる。この手法では、従来のCycleGANの損失に加え、図6のように、CycleGANによって得られる6種類の画像(sim, transferred sim, cycled sim, real, transferred real, cycled real)に対する物体検出のズレを損失(perception consistency loss)として導入している。

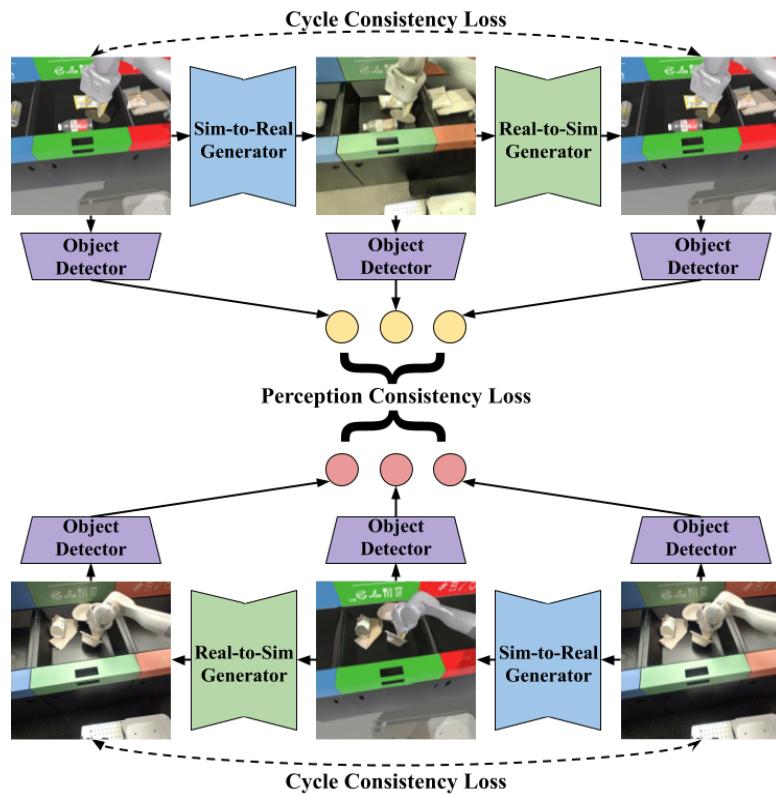


図6: RetinaGANの概要  
(文献[5]より転載)

RetinaGANの特徴を以下にまとめる。

- 物体情報を把握(object-aware)したsim2real変換ができる。
- タスクに固有せず、さまざまなタスクのための学習に利用できる。
- 模倣学習や教師あり学習ができるので、訓練しやすい。  
(cf. RL-CycleGANはRLモデルと連結して訓練させる必要がある。)
- ✗ bounding boxラベル付きの訓練データが必要である。

## 3. 考察

前節の文献調査をふまえて、建設現場のReal画像が少ない問題に対して、現時点では、RetinaGANが最も有効なアプローチであると考える。そのように考える理由は：

- RetinaGANが最新の研究(2021年6月発表)であり、従来のドメイン適応手法を上回る実験結果を示しているからだ。
- 建設機械でのAI活用において、現時点では、強化学習ベースの行動タスクに比べて、認識タスク(e.g. 人検知、障害物認識)に、より注力している段階であるからだ。
- RetinaGANの論文の権威性が十分高いと考えるからだ。
  - トップカンファレンス(ICRA)に採択されている。
  - 実績のある機関(Everyday Robots, Google, UC Berkeley)の著者たちによって発表されている。

建設機械／現場のためのAI研究開発で、RetinaGANを(調整して)採用することで、シミュレータから、Real画像を模擬したデータを大量に生成できるようになるはずである。

ただし、RetinaGANの論文内では、テストするReal環境にかなり近いシミュレータを用意していることが、懸念点である。そのため、特定の建設現場を模擬したシミュレータではなく、一般化された建設現場のシミュレータを用いて

も有効なのか検証が必要である。ただし、CycleGANは本来、2つのドメインの訓練データがペアになっている必要がないのが特徴であるため、ある程度大量のSimとRealのデータを用意すれば問題ないと予想している。

## 4. 結言

「建設現場のカメラ画像を大量に収集するのが容易ではない」という問題に対するアプローチを模索するために、文献調査を行い、それで得られた情報をもとに考察を行った。その文献調査では、少量のRealデータでの学習テクニックのうち、ドメイン適応に着目して調査を行った。2014年に発表されたGANの考え方をもとに、既存手法の問題点を解決するように、新たなドメイン適応手法が提案されてきた。そのなかで、2021年に発表されたRetinaGANは、既存の手法を上回る結果を示した。この調査結果をふまえて、現時点では、RetinaGANが最も有効なアプローチであると考察し結論づけた。そして、RetinaGANを的確に採用することができれば、シミュレータで得られた大量の画像を、現実世界の画像のようなデータに変換できるようになると予想している。ただし、建設現場のような複雑な環境に適用することができるかは不明であるため、今後も調査を続けたい。

## 謝辞

AIソリューション開発Gの川本駿さんより、建設現場のデータ収集の課題をご教授いただいたことが、本レポート作成のきっかけとなった。ここに篤く御礼申し上げる。

## 参考文献

1. [I.J. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville and Y. Bengio: Generative Adversarial Networks, Processing of Neural Information Systems 27 \(NIPS 2014\), 2014.](#)
2. [P. Isola, J. Zhu, T. Zhou and A.A. Efros: Image-To-Image Translation With Conditional Adversarial Networks, Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition \(CVPR\), 2017.](#)
3. [J. Zhu, T. Park, P. Isola, A.A. Efros: Unpaired Image-To-Image Translation Using Cycle-Consistent Adversarial Networks, Proceedings of the IEEE International Conference on Computer Vision \(ICCV\), 2017.](#)
4. [K. Rao, C. Harris, A. Irpan, S. Levine, J. Ibarz and M. Khansari: RL-CycleGAN: Reinforcement Learning Aware Simulation-to-Real, Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition \(CVPR\), 2020.](#)
5. [D. Ho, K. Rao, Z. Xu, E. Jang, M. Khansari, Y. Bai: RetinaGAN: An Object-aware Approach to Sim-to-Real Transfer, Processing of International Conference on Robotics and Automation \(ICRA\) 2021, 2021.](#)