

# Sprint16 論文紹介

# Generative Adversarial Nets

2014

Ian J. Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair,  
Aaron Courville, Yoshua Bengio

<https://arxiv.org/pdf/1406.2661.pdf>

## どんなもの？

- 生成モデルの Generator と識別モデルの Discriminator を敵対させるプロセスを介した、新しく提案されたフレームワーク
- Facebook の AI 研究所所長である ヤン・ルカン 氏によると、「機械学習において、この 10 年間で もっとも おもしろい アイデア」と形容された

## どうやって有効だと検証した？

- MNIST, TFD(Toronto Face Database) をデータセット
- DBN, Stacked CAE, Deep GSN のモデルと対数尤度を比較し有効性を研究した

## 技術の手法や肝は？

- Discriminator から流れる勾配によってのみパラメータが更新される点がモデルの他の拡張性に寄与している

## 議論はある？

- 欠点として、Generator の確率分布に明示的な表現方法がないこと、学習中に Generator と Discriminator が適切に同期する必要があることがあげられてい
- 利点として、学習中の推論が不要であること、様々な関数をモデルに組み込むことが可能であることがあげられている

## 先行研究と比べて何がすごい？

- DBN(Deep belief networks)は、単一の無行層と複数の有行層を含んでおり、計算上の問題があった
- GSN(generative stochastic network)は、マルコフ連鎖の 1 つのステップを実行するものと見て取れるが、GAN には生成中のフィードバックを必要としないため、マルコフ連鎖を使わないでモデル構築を行っている

## 次に読むべき論文は？

- GAN は色々なモデルに組み込みが可能という利点があるとのことで、その後発展した GAN の論文を読む

# Conditional Generative Adversarial Nets

2014

Mehdi Mirza, Simon Osindero

<https://arxiv.org/pdf/1411.1784.pdf>

## どんなもの？

- GeneratorとDiscriminatorの両方に条件付けを行い、データを供給するだけで構築できるGANの紹介
- この論文ではクラスラベルを条件とするMNISTの数値を生成することを示す

## 議論はある？

- 現在の実験では、各タグを個別に使用するしているだけであるが、複数のタグを同時に使用することで（「セット生成」の1つとして生成上の問題を効果的に提起する）、より良い結果が達成できると想定している

## どうやって有効だと検証した？

- MNIST, UGM(user-generated metadata)の2つのデータを使って、DBN, StackedCAE, DeepGSN, Adversarial netsのモデルと比較し検証した

## 先行研究と比べて何がすごい？

- 自然言語コーパスを使用して、意味のあるラベルのベクトル表現を学習する際、見学習の言語でも「近い」ことが多いという（たとえば、「椅子」ではなく「テーブル」を予測する）ことでラベル付の意味がある
- 画像認識の分野でも、アノテーションを行うことで1枚の画像から複数ラベル付を行うことができている

## 技術の手法や肝は？

- Generator, Discriminatorともに、クラスラベルもinputのデータとして条件付を行っている

## 次に読むべき論文は？

- GANは色々なモデルに組み込みが可能という利点があるとのことで、その後発展したGANの論文を読む

# Unpaired Image-to-Image Translation using Cycle-Consistent Adversarial Networks

2017

Jun-Yan Zhu, Taesung Park, Phillip Isola, Alexei A. Efros

<https://arxiv.org/pdf/1703.10593>

## どんなもの？

- ソースドメインからターゲットドメインへの変換タスク（ある風景写真をモネ風のイメージへ）
- 多くの場合、pairとなる画像イメージを入手することができないため、「循環による一貫性を持つ敵対性ネットワーク」を用いた「教師なし学習」によって画像変換タスクのモデルを構築

## 議論はある？

- 色を変換するタスク（馬をシマウマに変換する）などは成功していたが、ジオメトリを変形するタスクでは失敗（犬を猫に変換する）
- 写真からラベルに変換するタスクにおいては、木と建物を混同してしまうといったこともあり、曖昧さを解決するためには、何らかの弱い制約が必要だと考えられている

## どうやって有効だと検証した？

- 次の5つのモデルと比較して有効性を検証
- CoGAN, SimGAN, Feature loss + GAN, BiGAN / ALI, pix2pix

## 先行研究と比べて何がすごい？

- 写真のテイストを変えたり、物体を変身させたり、季節の遷移させたり、画像のエンハンスメントなど、多くの事例で成果をあげている

## 技術の手法や肝は？

- ソースドメイン→ターゲットドメインへの敵対性損失（従来のGANと同様）
- ターゲットドメイン→ソースドメインへの敵対性損失（逆も行う）
- 上記に一貫性を持たせるために、「変換後のイメージから入力データを復元したイメージ」と「元の入力データ」を比較するサイクル一貫性損失も導入してモデル学習を行った

## 次に読むべき論文は？

- GANは色々なモデルに組み込みが可能という利点があるとのことで、その後発展したGANの論文を読む

# Few-shot Knowledge Transfer for Fine-grained Cartoon Face Generation

2020

Nan Zhuang, Cheng Yang  
<https://arxiv.org/pdf/2007.13332>

## どんなもの？

- ソースドメインからターゲットドメインへの変換タスク（ある風景写真をモネ風のイメージへ）
- 多くの場合、pairとなる画像イメージを入手することができないため、「循環による一貫性を持つ敵対性ネットワーク」を用いた「教師なし学習」によって画像変換タスクのモデルを構築

## どうやって有効だと検証した？

- 基本となるモデルと、そのモデルから重みを転移させた数パターンでの生成結果を比較し、重みを転移させたことの有効性を検証した

## 技術の手法や肝は？

- 比較的多くのリアル写真と漫画画像が入手できるグループを軸として、Generatorを学習
- 他のグループでも漫画のタッチは同一なので、Generatorの中間部分を他のグループと共有として、入力層と出力層を個別に学習させていくことで、少ないイメージしかないグループでも特徴を抽出することに成功した

## 議論はある？

- 1グループ数ショットの画像でグループの特徴を抽出し、画像生成を試みているが、少しぼやけている生成結果もある
- グループ間の違いを明示的に捉えるために新しいアテンションブロックを導入することで改善を図ることを検討できる

## 先行研究と比べて何がすごい？

- 少ない画像しか入手できないセグメントであってもそのセグメントの特徴をうまく反映させてアニメ画像の生成に成功している

## 次に読むべき論文は？

- 本論文は顔写真のみにフォーカスしているが、画像全体をアニメ風にする研究もあり、アルゴリズムの違いを把握したい（AnimeGAN）

# U-Net: Convolutional Networks for Biomedical Image Segmentation

2015

Olaf Ronneberger, Philipp Fischer, Thomas Brox  
<https://arxiv.org/abs/1505.04597>

## どんなもの？

- ネットワーク構造はコンテキストを掴むための収縮パスと正確なローカライゼーションを可能にする収縮パスに対称的な構造を持つ拡張パスによって構成されている
- U-Netのネットワークがとても少ない訓練画像からend-to-endで学習し、ISBIチャレンジにおけるセグメンテーションにおいて従来のベストな手法（sliding-window convolutional network）を上回ったことを示している

## どうやって有効だと検証した？

- 3つの異なるセグメンテーションタスクで従来の手法と比較して検証

## 技術の手法や肝は？

- downsamplingとupsamplingが対称的に構成されており、downsamplingにおいては画像のサイズが縦横それぞれ半分になる際にチャンネルの数を倍に増やしている
- upsamplingの際は対称となるdownsamplingのパスから切り出した(cropped)画像を足し合わせている。画像を切り出す理由は、畳み込み処理の際にパディングを行わないことで画像が小さくなってしまいうのが理由

## 議論はある？

- 特になし

## 先行研究と比べて何がすごい？

- u-netアーキテクチャは、生物医学で非常に優れたパフォーマンスを実現した。

## 次に読むべき論文は？

- 画像セグメンテーションで直近の研究論文を確認