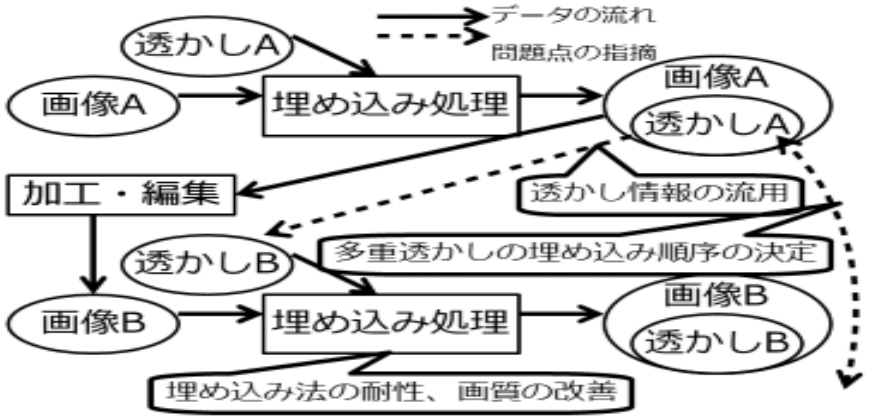
CNNの中間層とVAEの電子透かしへの応用

木下研究室　201403762　牛山　尚織

研究動向と問題点

- 一般的な電子透かしでは、透かし情報をコピーされてほかのコンテンツに利用される可能性がある。
- また、二次利用による多重の電子透かしの優先順位や加工内容との加工関係を明確にできない。



目的

- ＣＮＮの中間層を用いた知覚ハッシュの構成法
- 2件前の発表(小藤田)

  中間層の位相幾何学的構造→知覚ハッシュ値
- 1件前の発表(小澤)

  中間層のプーリング層の512個のノードを主成分分析→知覚ハッシュ値

- 本研究
  - 画像の中間層を用いた、VAE/GANによる知覚ハッシュ値の生成

研究の概要(中間層情報の生成)

画像の特徴を抽出するために中間層の情報を利用
Signal：知覚ハッシュ値を求める対象の画像の集合（100枚）
Noise：Signal以外の画像の集合（200枚）

画像の変形による学習データオーグメンテーション（keras）

(原画像から複数の変形画像を作成)

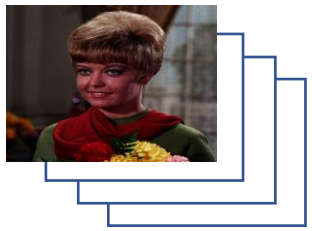
Signalとしての画像

入力

kerasの学習済みモデルVGG16を用いたFine-Tuning

出力

カテゴリSignal



Noiseとしての画像

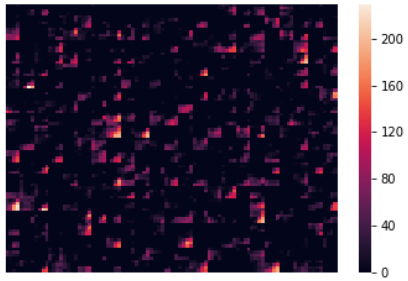
入力

Convolutional Neural Network

出力

カテゴリNoise

中間層の情報を抽出



Signalの画像を入力したとき、出力がクラスSignalとなり、Noiseの画像を入力したときには、出力がクラスNoiseになるように学習させる

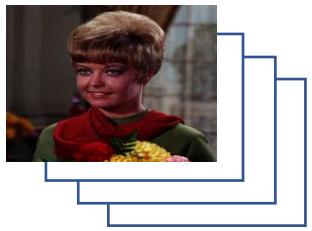
研究の概要(データオーグメンテーション)

画像の特徴を抽出するために中間層の情報を利用
Signal：知覚ハッシュ値を求める対象の画像の集合（100枚）
Noise：Signal以外の画像の集合（200枚）

画像の変形による学習データオーグメンテーション（keras）

(原画像から複数の変形画像を作成)

Signalとしての画像



Noiseとしての画像

入力

入力

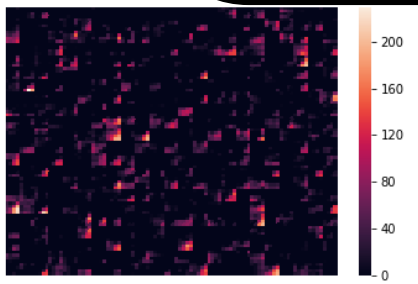
kerasの

デル

Conv

Neu

・Jupyter Notebook上で画像処理ライブラリであるOpenCVを使い、回転、移動、拡大、縮小させた画像を用意する。



Signalの画像を入力したとき、出力がクラスSignalとなり、Noiseの画像を入力したときには、出力がクラスNoiseになるように学習させる

研究の概要(データオーグメンテーション)

画像の特徴を抽出するために中間層の情報を利用
Signal：知覚ハッシュ値を求める対象の画像の集合（100枚）
Noise：Signal以外の画像の集合（200枚）

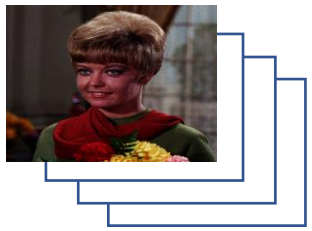
画像の変形による学習データオーグメンテーション（keras）

(原画像から複数の変形画像を作成)

Signalとしての画像

入力

kerasの
デルV
たFi



Noiseとしての画像

入力

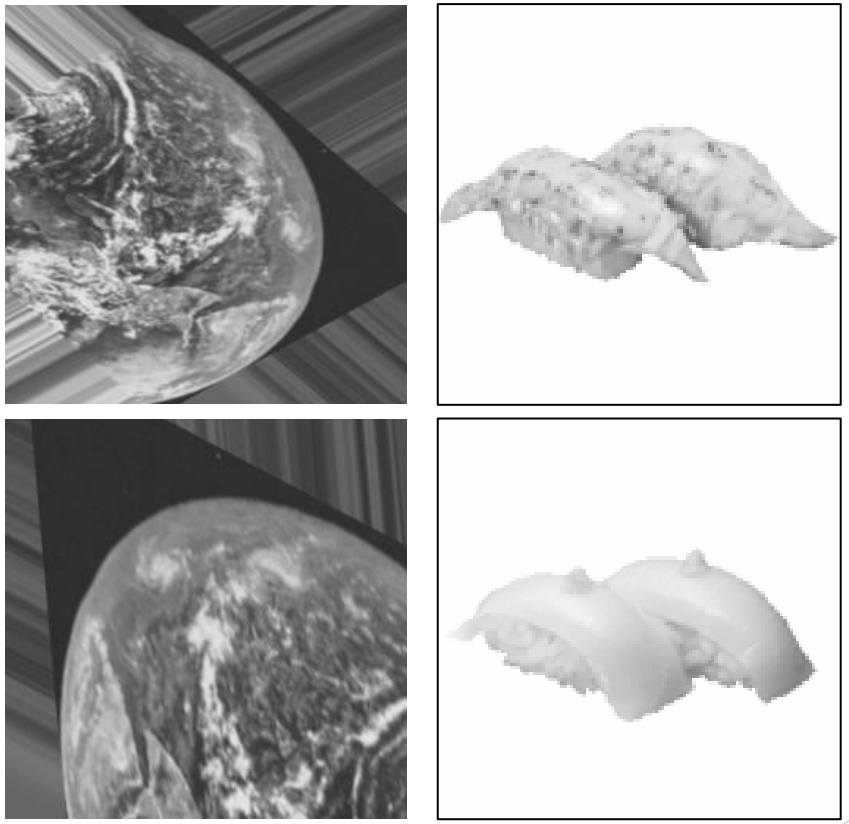
Neu

Noiseになるように学習させる

・本実験では、Signalとしての原画像から、加工・編集を行った画像100枚ずつ
・その他の画像として、違う種類の画像を加工・編集を行った画像と、複数の類似画像100枚ずつ

提案手法
(データオーグメンテーション2)

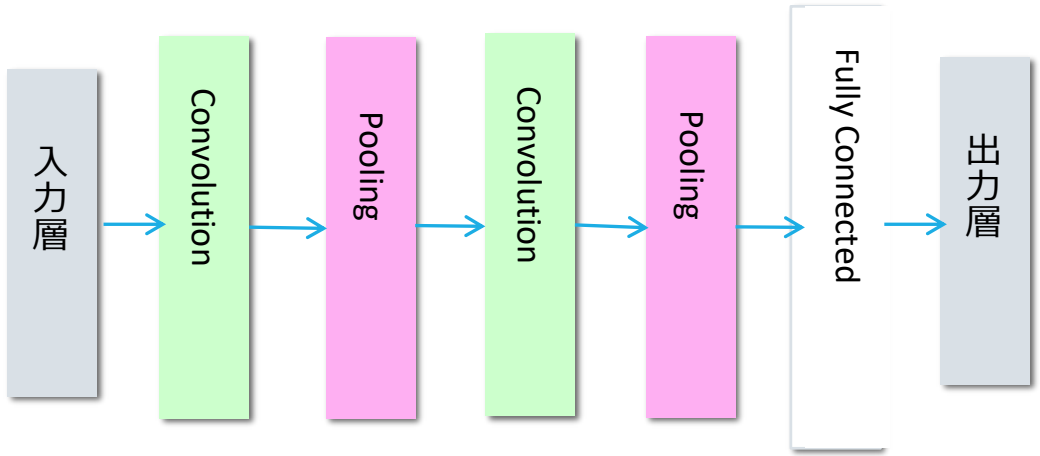
Signalとしての画像集合
（girl）



Noiseとしての変形画像と複数の類似画像
（earthと寿司）

畳み込みニューラルネットワーク(CNN)

- 畳み込みニューラルネットワーク(CNN)は、「畳み込み層」や「プーリング層」などのいくつかの層をもつニューラルネットワークで、画像認識の分野で特に優れた性能を発揮しています
- 画像内の特徴量を圧縮することで、画像の変形などに強くすることで計算量を下げることができる



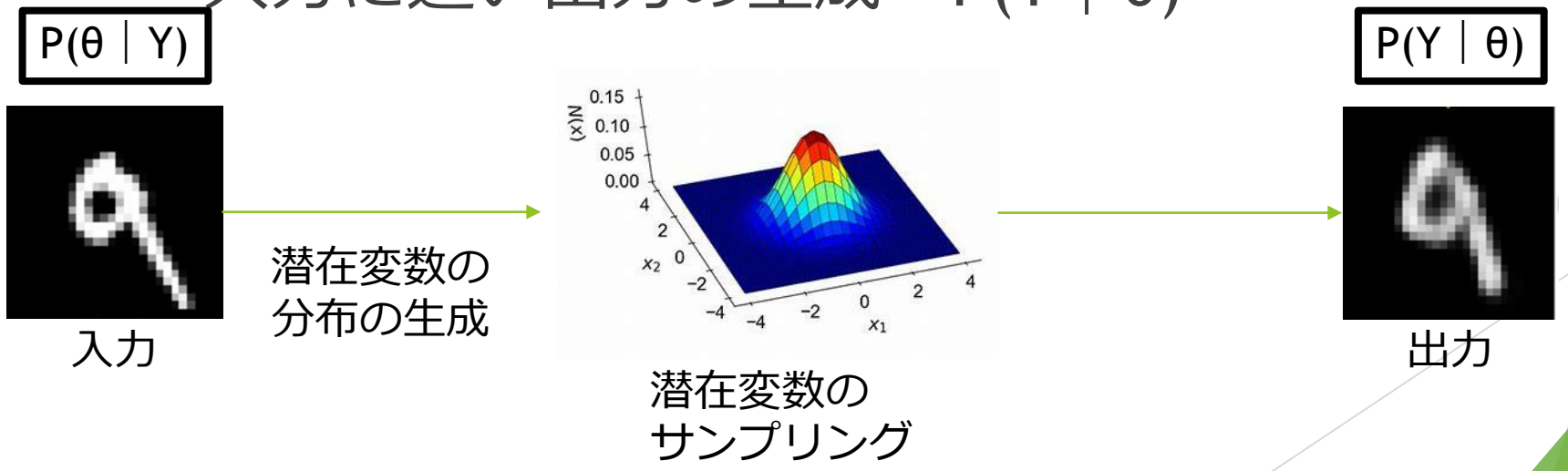
VAE(Variational Auto Encoder)

- 入力→潜在変数の分布の生成　P(θ｜Y)

  θ:潜在変数

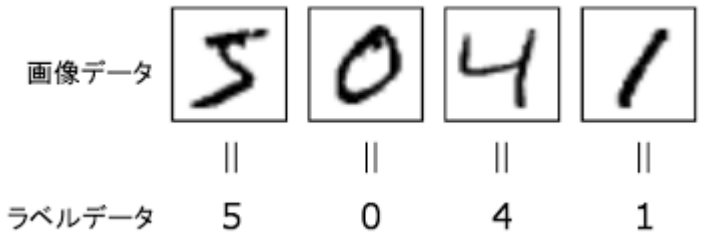
- 潜在変数からのサンプリング

  →入力に近い出力の生成　P(Y｜θ)



MNIST
(Mixed National Institute of Standards and Technology database)

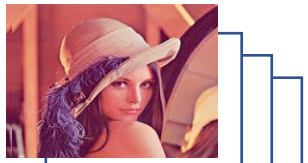
- MNIST(Mixed National Institute of Standards and Technology database)とは、手書き数字画像60,000枚と、テスト画像10,000枚を集めた、画像データセット
- 画像データと画像に書かれた数字の正解となるラベルデータで構成されている



GAN
(Generative Adversarial Networks)

- GANは、敵対的生成ネットワークのことで、正解データを与えることなく特徴を学習する生成モデル
- データから特徴を学習することで、実在しないデータを生成したり、存在するデータの特徴に沿って変換できる

研究の概要(中間層情報のVAE/GANによる解析)



入力

Signalとしての画像

Fine-Tuning済みCNN

中間層情報

入力

Noiseとしての画像

Fine-Tuning済みCNN

中間層情報

VAE

VAEによる出力

GAN

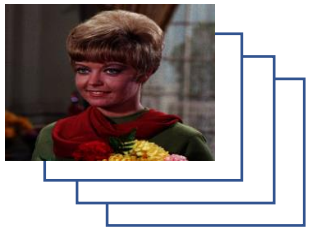
GANによる敵対的データから弁別を学習

知覚ハッシュ値の生成
（加工によっても変化しない値）

研究の概要(VAEのみによる解析)

Signalとしての画像

入力



Noiseとしての画像

入力

VAE

VAEによる
出力

知覚ハッシュ値の生成
（加工によっても変化しない値）

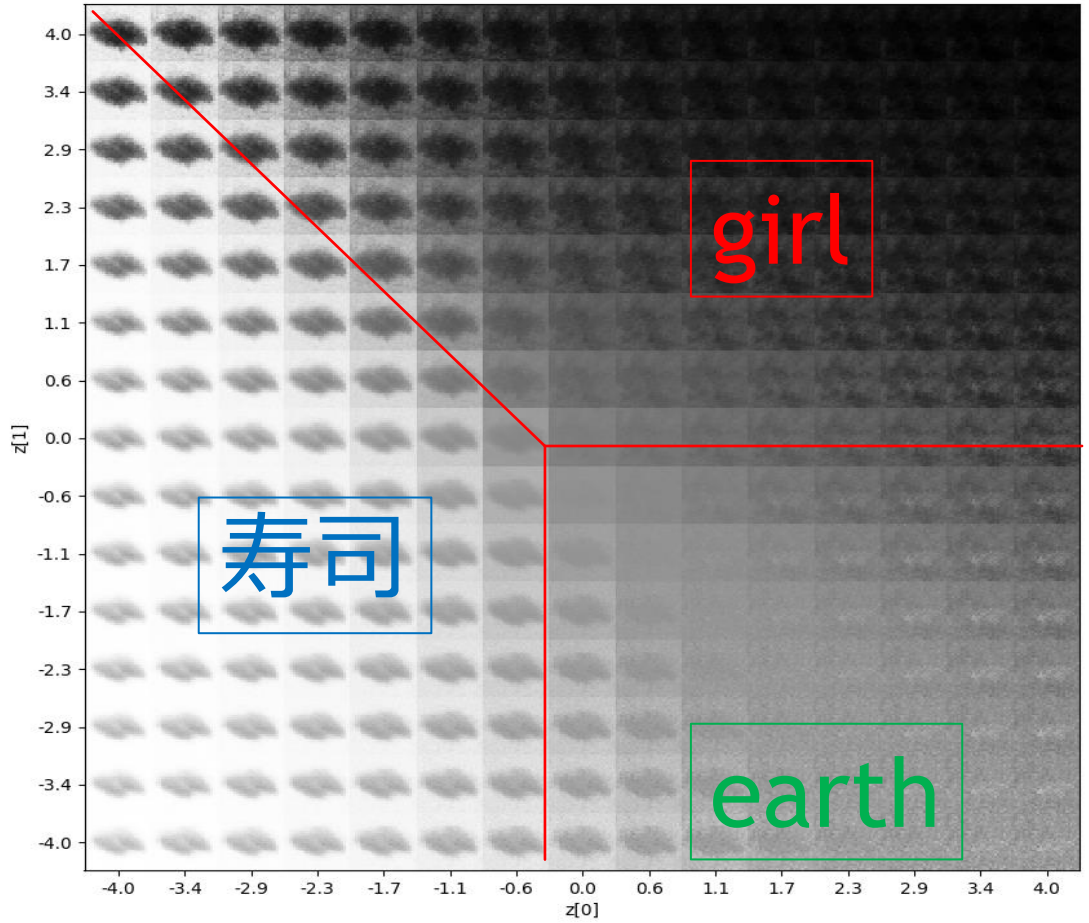
・訓練データを元に潜在変数の出力

VAEの実装

VAEのMNISTによる、画像生成のサンプルプログラムからSignalとNoiseの画像を読み込み、潜在変数の分布を出力する

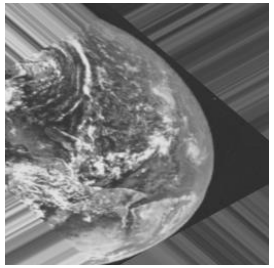
対象が画像情報なので、潜在変数は2次元で説明される

出力画像のクラスタリングされた2次元マップ



VAEは通常このように画像が分類されて、
それぞれの画像集合の特徴は
2次元マップのクラスター上に
潜在変数Ｚとして分布している

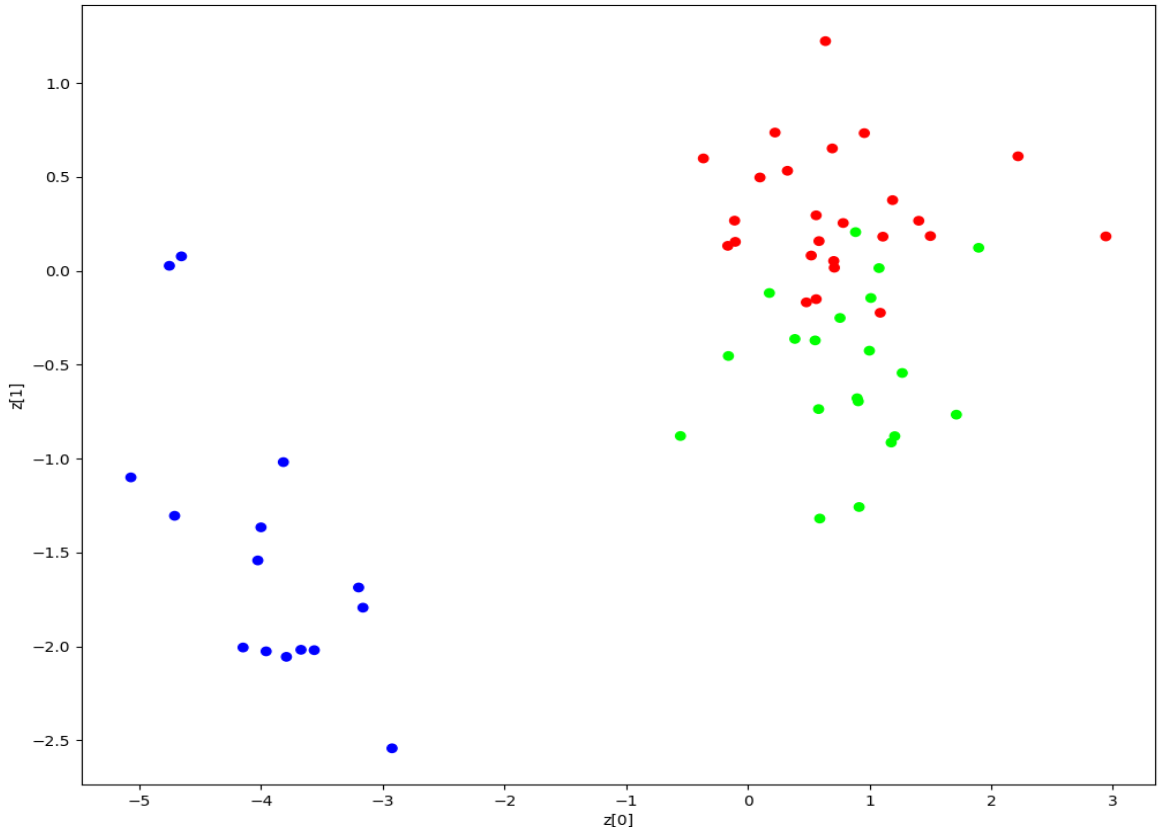
girl



earth

寿司

潜在変数Zの分布



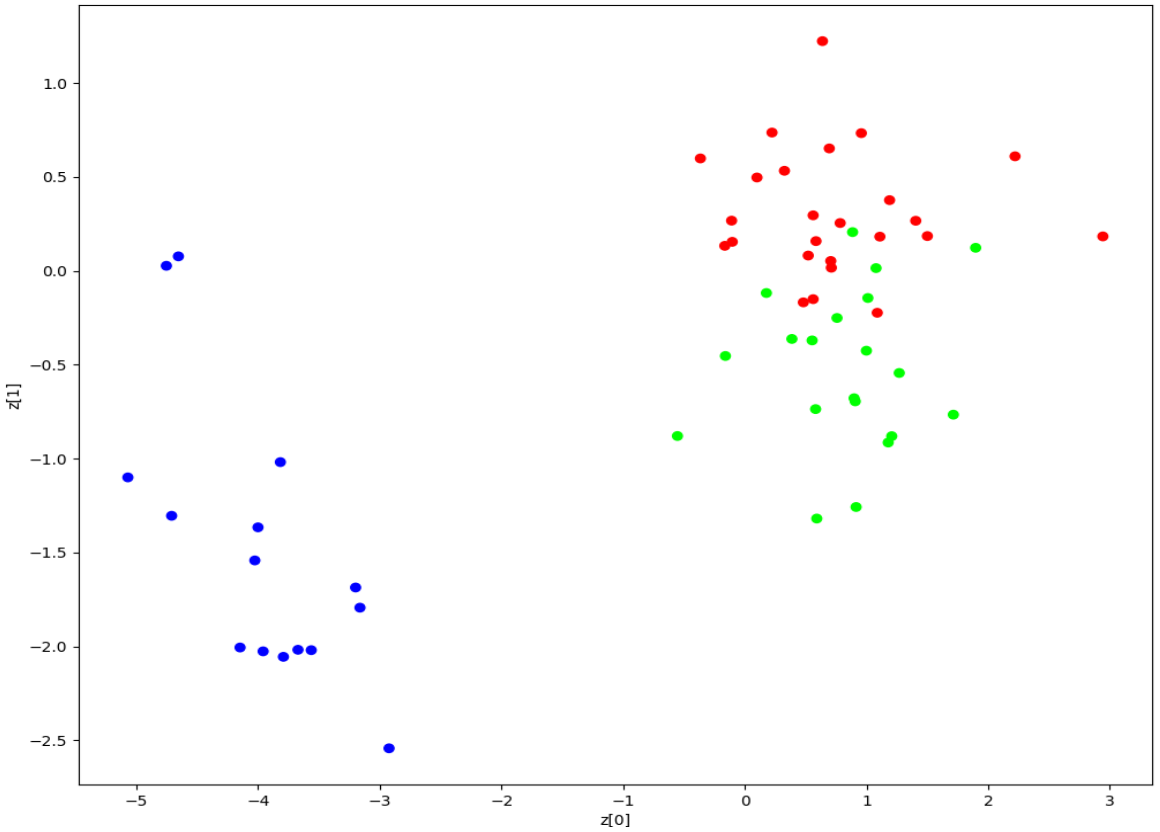
3つの画像の**潜在的確率変数Z**の**２次元マップがVAE出力の一つである**

赤点：Signalの集合

緑点：Noiseの集合１

青点：Noiseの集合２

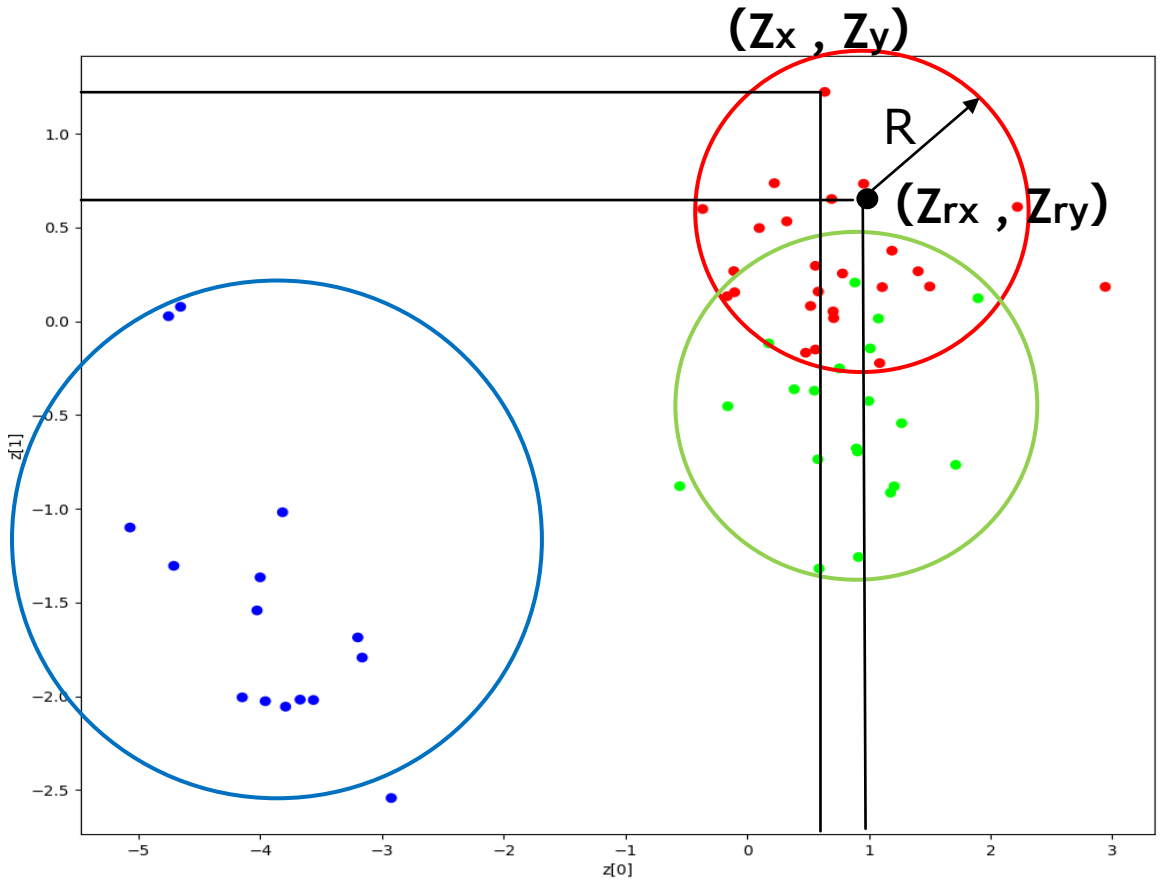
潜在変数Ｚの分布



赤と緑は、それぞれ同一の画像を加工・編集した画像であり、一定の範囲に集中して分布している

この分布が、加工・編集に影響されない画像の特徴と考えられる

知覚ハッシュ値の導出



赤点：Signalの集合
緑点：Noiseの集合１
青点：Noiseの集合２

2次元潜在変数の分布から重心を取り、知覚ハッシュ値として値をとる

$$重心(知覚ハッシュ) = \left(\frac{\sum Zrx}{N} , \frac{\sum Zry}{N}\right)$$

$$\sqrt{\left(Zx - \frac{\sum Zrx}{N}\right)^2 + \left(Zy - \frac{\sum Zry}{N}\right)^2} \leq R$$

まとめ

- VAEに画像を直接入力することだけで、出力した2次元潜在変数の分布から重心を取ることで、知覚ハッシュ値の生成をすることができた

- 今後の課題としては、実験結果より2次元潜在変数の分布が重複しているため、GANの敵対的学習により厳密に弁別できるように学習させることがあげられる