

VAE に基づく距離画像の欠損補間及び物体認識への応用

○本田 圭†, 三輪 顕太郎†, 山内 悠嗣†

†: 中部大学

er17064-1155@sti.chubu.ac.jp, tr20012-5568@sti.chubu.ac.jp, yuu@isc.chubu.ac.jp

概要: センサから物体までの距離を測定可能な距離センサが普及しており, 周辺環境の認識等のセンシング技術に利用されている. しかし, 距離センサは物体の材質や形状によって距離が計測できない問題がある. そこで, 本稿では変分オートエンコーダ(VAE)を用いた距離値欠損の補間手法を提案する. 欠損のないシミュレーション画像を用いて学習することで学習データと類似した距離画像を出力する. 実験の結果, 提案手法により距離値の欠損の補間が可能であり, 補間した距離画像を用いた物体認識の精度が向上することを確認した.

<キーワード> 欠損補間, 変分オートエンコーダ, 距離画像

1. はじめに

近年, センシング技術の発達により三次元構造を計測可能な距離センサが物体認識やロボットの自己位置推定等の技術に利用されている. 距離センサは距離値の計測方法として Time of Flight(ToF)方式やパターン照射方式が採用されている. 例えば, ToF 方式はセンサの発行部から赤外光を照射し, 物体に反射した赤外光を受光部で観測する. そして, 赤外光を受光部で観測するまでの光の飛行時間から距離を算出する. しかし, 光を透過・吸収・反射する物体や奥行のある物体の距離を正確に計測できず, 図 1 のような欠損や誤差が生じる問題がある. このような欠損を含んだ距離画像を用いて物体認識や自己位置推定を行うと性能が低下する恐れがある.



図 1. 距離画像の欠損例.

このような問題に対して, ローカルフィルタをベースとした手法[1,2]やマカロフ画像合成を用いた手法[3]により距離値の欠損を補間する手法が提案されている. 近年では深層学習を用いて欠損を補間する手法[4,5,6,7,8,9]が提案されている. し

かし, これら深層学習に基づく手法は学習のために大量のカラー画像と距離画像が必要であり, 実世界でこれらの画像を収集することはコストが生じる. また, 距離画像は欠損を含むため, 学習に使用すると悪影響を及ぼす可能性がある.

本稿では変分オートエンコーダ(VAE)[10]を用いて距離画像の欠損を補間する方法を提案する. 欠損を含まない距離画像のみを用いて VAE を学習することで, 学習データと類似した距離画像を VAE により出力することが可能となる. 提案手法ではこの特性を利用し, VAE に欠損を含んだ距離画像を入力することで, 学習に用いた画像に近い画像として距離を補間する. また, 提案手法は学習データをセンサシミュレーションで獲得する. これにより, 低コストで自動的かつ大量に学習データの収集が可能となる.

2. 関連研究

2.1. 距離補間

距離画像の欠損を穴埋めする手法が研究されている. Junho らはラプラシアンピラミッドを基にした畳み込みニューラルネットワークにより距離画像の欠損を補間する手法[4]を提案した. Zhang らは, カラー画像から推定した幾何学情報を利用して深層学習により距離画像の欠損を補間する手法[5]を提案した. これらの欠損の穴埋め手法は学習データとして欠損画像と ground truth が必要となる. しかし, 実世界の距離画像は欠損を含むため, ground truth の作成が必要となる. 最近の研究では強化学習による距離画像を入力とした点群ベースの欠損補間手法[6]が提案されている. この研究は

実世界で得られない ground truth を 3D モデルから獲得している。

2.2. 距離補間

距離画像の欠損を補完する方法の他にも、他の情報から距離を推定する研究が報告されている。カラー画像から深層学習により距離を推定する手法[8,9]が提案されている。また、Xie らはマルチスケールの特徴マップを融合させる手法[9]を提案した。これらの手法は実世界でセンシングしたデータに基づき推定器を学習するが、大量のデータを必要とする問題が発生する。

2.3. 提案手法の概要

本稿では変分オートエンコーダ(VAE)に基づいた欠損補間法を提案する。提案手法の特長は次の2点である。

- (1) VAE による距離画像の欠損補間
VAE は学習データに含まれない異常な画像がネットワークに入力されると、学習データに近い画像として復元される。本研究ではこの特性を距離画像の欠損補間に利用する。
- (2) シミュレーションによるデータ自動収集
VAE の学習には欠損を含まない大量の距離画像が必要である。実世界で欠損を含まない画像を収集することが難しいため、本研究ではセンサシミュレーションにより学習用画像を自動的に生成する。

3. 提案手法

オートエンコーダ(AE)[11]では、エンコーダ(E)により入力画像を低次元なベクトルに圧縮し、そのベクトルをデコーダ(D)により画像を復元する。学習データに含まれない異常な画像が入力されると、 E と D の処理過程で異常な特徴が欠落し、学習データに近い画像として復元される。提案手法では、この AE の復元能力を利用して距離画像の欠損を補間する。

提案手法の流れを図 2 に示す。学習プロセスでは、シミュレータを利用して生成した欠損のない距離画像を用いて E と D を学習する。なお、本研究では敵対的学習を導入することで鮮明な画像を生成可能とした Intropective Variational Autoencoders(IntroVAE)[12]を導入する。欠損補間プロセスでは、学習したネットワークに欠損を含む距離画像

を入力し、欠損を補間した複数の距離画像を生成する。そして、複数の候補画像から欠損領域外で最も整合性が高い画像を補間画像として出力する。

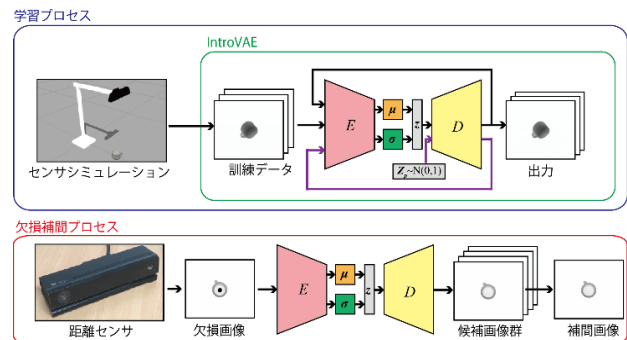


図 2. 提案手法の流れ。

3.1. 変分オートエンコーダ(VAE)

AE では入力データを E により潜在変数 z として低次元なベクトルに圧縮していたが、VAE は畳み込みニューラルネットワークで構成され、潜在変数を正規分布(平均 μ , 分散 σ)と仮定して圧縮する。VAE は高解像度画像を生成すると画像が不鮮明になる問題を抱えているため、本研究では鮮明な画像を生成可能な IntroVAE を導入する。

IntroVAE は VAE に敵対的学習を導入することで鮮明な画像を生成可能としている。VAE の E を識別器、 D を生成器と見立てて敵対的学習を行う。

E は入力画像を真のデータ、 D による復元画像と生成画像を偽のデータとし、データの真偽を識別するように学習する。そのため、入力画像 X 及び、 D の復元画像 X_r 及び、乱数で生成した潜在変数 z_p から獲得した生成画像を E に入力する。 E の損失関数 L_E は式(1)により算出する。 D は復元画像と生成画像を入力画像と類似したデータとなるように学習する。 D の損失関数 L_D は式(2)で示される。

$$L_E = L_{REG}(E(X)) + \alpha \sum_{s=r,p} [m - L_{REG}(E(D(z_s)))]^+ + \beta L_{AE}(X, X_r) \quad (1)$$

$$L_D = \alpha \sum_{s=r,p} L_{REG}(E(D(z_s))) + \beta L_{AE}(X, X_r) \quad (2)$$

再構成誤差 L_{REG} は式(3)を用いて計算され、 μ と σ を正規分布に制約する。復元誤差 L_{AE} は式(4)のように二乗和誤差を算出する。

$$L_{REG} = \frac{1}{2} \sum_{j=1}^J (1 + \log((\sigma_j)^2) - \mu_j^2 - (\sigma_j)^2) \quad (3)$$

$$L_{AE} = \frac{1}{2} \|X_r - X\|^2 \quad (4)$$

3.2. Reparameterization Trick

VAE は畳み込みニューラルネットワークで構成され、誤差逆伝播法により重みを学習する。しかし、VAE のように潜在変数を正規分布と仮定すると誤差逆伝播法を適用することができない。そのため、一般的には Reparameterization Trick を導入することによりこの問題を解決している。

正規分布に従った乱数 ϵ 及び、 E から出力された平均 μ と分散 σ を用いて、式(5)により潜在変数 z に再パラメータ化する。

$$z = \mu + \sigma * \epsilon \quad (5)$$

これにより誤差逆伝播法を適用可能としている。提案手法では欠損を含んだ距離画像をVAEにより補間するが、乱数 ϵ の影響を強く受けた補間画像が生成される。図 3 に乱数による候補画像の例を示す。乱数によっては物体の形状や姿勢が変化する問題がある。

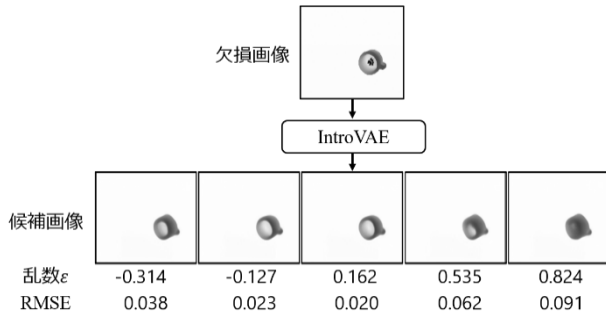


図 3. 乱数による候補画像の例.

3.3. 欠損補間画像の選択

欠損を含む距離画像の出力画像が不安定になる問題を解決するため、多数の乱数から複数の距離画像 \hat{X} を生成し、欠損領域外で最も整合性が高い画像を補間画像とする。

入力した欠損画像 X_{in} と候補画像 $\hat{X} \in \hat{X}$ を用いて式(6)より求めた平均平方二乗和誤差(RMSE) が小さい候補画像を補間画像 \hat{X}_{out} として採用する。

$$\hat{X}_{out} = \arg \min_{\hat{X} \in \hat{X}} \sqrt{(X_{in} - \hat{X})^2} \quad (6)$$

この時、距離画像の欠損領域は距離値の欠落により画素値 0 として表現されているため、欠損以外の領域において最も整合性がある画像を出力することになる。

4. 学習データの収集

提案手法ではVAEを学習するために欠損を含まない大量の距離画像が必要である。しかしながら、

実空間においてセンサにより距離画像を収集すると、撮影環境や対象物体によっては計測した距離に欠損や誤差が生じる。そこで、本研究ではセンサシミュレーションにより欠損を含まない距離画像を自動獲得する。センサシミュレーションを行うために、ロボット分野で利用されるシミュレータ Gazebo[13] を利用する。Gazebo では OGRE(Object-Oriented Graphics Rendering Engine)[14] を採用しており、高画質な 3D グラフィックスを描画することが可能である。

学習データを収集するために、Intel RealSense D415 と 3D モデリングソフト RecFusion により実験に使用する 5 アイテムの 3D モデルを作成した。Gazebo の仮想空間に固定カメラを設置し、3D モデルの位置と姿勢を変えながら配置することで対象物体を自動的に撮影することができる。図 4(a) に対象物体、(b) に作成した 3D モデル、(c) にセンサシミュレーションにより撮影した距離画像を示す。



図 4. 各アイテム及び 3D モデルとセンサシミュレーションにより取得した距離画像.

5. 評価実験

評価実験により提案手法の有効性を確認する。

5.1. データセット

実験には距離画像に欠損が発生しやすい cup, bottle, spray, scissors, strainer の 5 アイテムを使用する。各アイテムの 3D モデルを作成し、Gazebo を用いたセンサシミュレーションにより自動的に収集した 54,807 枚の 212×256 画素の距離画像を実験に使用する。54,307 枚を学習に使用し、残りを評価用画像とした。補間精度の比較のために評価指標として RMSE を用いる。

5.2. ネットワーク構成

学習に使用した IntroVAE の E と D はそれぞれ 8 層の畳み込みネットワークで構成し、潜在変数は 256 次元、ハイパーパラメータは $\alpha = 0.25, \beta = 25, m = 180$ とした。最適化手法は Adam を使用し、バッチサイズ 50 で学習した。また、欠損補間プロセスでは 4,000 個の乱数により候補画像を探索した。

5.3. 距離画像における欠損補間の比較

VAE 及び提案手法により距離画像の欠損を補間した結果を表 1 に示す。なお、表 1 における欠損画像は、欠損を与える前後の画像から求めた RMSE を意味する。VAE の誤差は欠損画像と比べて 22.2% 低下、提案手法では誤差が欠損画像と比べて 48.1% 低下した。これにより、提案手法が距離画像の欠損をより補間できることを確認した。

表 1 欠損補間の誤差の比較。

	cup	bottle	spray	scissors	strainer	平均
欠損画像	0.058	0.057	0.067	0.027	0.059	0.054
VAE	0.062	0.039	0.046	0.020	0.045	0.042
提案手法	0.050	0.015	0.033	0.014	0.028	0.028

提案手法の補間結果を図 6 に示す。図 5.(a), (b), (d), (e) に着目すると欠損が補間されていることが確認できる。図 5.(c) では、補間はできているが物体の姿勢が変化し、誤差が増大していることを確認できる。

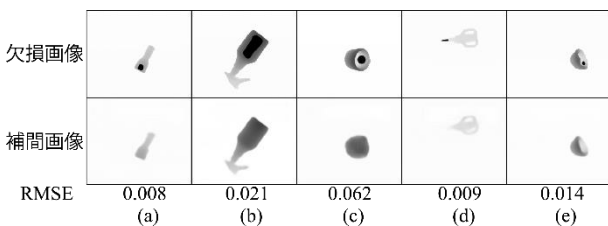


図 5. 欠損画像と提案手法により補間した画像の例。

5.4. 物体認識への応用

欠損を補間した画像に対して物体検出手法である Single Shot Multibox Detector(SSD)[15]を適用し、距離画像の欠損補間が物体認識に対して有効であるか確認する。学習には 54,307 枚の距離画像と廃毛画像として Pacal VOC データセットの画像 3,000 枚、実機センサの距離画像 1,000 枚を使用した。評価指標として Average Precision(AP)を用いる。検出結果を表 2 に示す。提案手法は欠損画像と比べて検出精度は向上していることが確認できる。特に欠損領域が大きい bottle において高い効果があることがわかる。図 7 に検出結果の例を示す。欠損

画像と比較すると提案手法では精度が向上していることが確認できる。

表 2 SSD による物体検出結果の比較

	cup	bottle	spray	scissors	strainer	mAP
欠損画像	0.6070	0.0002	0.9848	0.7312	0.0932	0.4833
VAE	0.7960	0.0909	0.9991	0.4956	0.6976	0.6158
提案手法	0.7989	0.5847	0.9991	0.6091	0.8967	0.7777

6. おわりに

本稿では変分オートエンコーダによる距離画像の欠損補間を提案した。欠損のないシミュレーション画像を用いて学習することで、欠損画像から学習データと類似した距離画像を生成可能とした。評価実験により提案手法は距離画像の欠損が補間可能なことを確認した。また、提案手法により距離画像の欠損に起因した物体認識精度の低下を改善可能なことを確認した。

参考文献

- [1] Xiaoong Shen et al. : Mutual-Structure for Joint Filtering, ICCV, 2015.
- [2] L. Zhang et al. : Depth enhancement with improved exemplar-based inpainting and joint trilateral guided filtering, ICIP, 2016.
- [3] David Doria et al. : Filling large holes in lidar data by inpainting depth gradients, IEEE, 2012.
- [4] Junho Jeon et al. : Reconstruction-based Pairwise Depth Dataset for Depth Image Enhancement Using CNN, ECCV, 2018.
- [5] Yinda Zhang et al. : Deep Depth Completion of a Single RGB-D Image, CVPR, 2018.
- [6] Xiaoguang Han et al. : Deep Reinforcement Learning of Volume-guided Progressive View Inpainting for 3D Point Scene Completion from a Single Depth Image, CVPR, 2019.
- [7] D. Eigen et al. : Depth map prediction from a single image using a multiscale deep network, NIPS, pages, 2366–2374, 2014.
- [8] David Eigen et al. : Predicting Depth, Surface Normals and Semantic Labels with a Common Multi-Scale Convolutional Architecture, IEEE, 2015.
- [9] Junyuan Xie et al. : Deep3D: Fully Automatic 2D-to-3D Video Conversion with Deep Convolutional Neural Networks, ECCV, 2016.
- [10] Diederik P. Kingma et al. : Auto-Encoding Variational Bayes, ICLR, 2014.
- [11] Hinton, Geoffrey E et al. : Reducing the dimensionality of data with neural networks, science 313.5786 (2006): 504–507.
- [12] Huaibo Huang et al. : IntroVAE: Introspective Variational

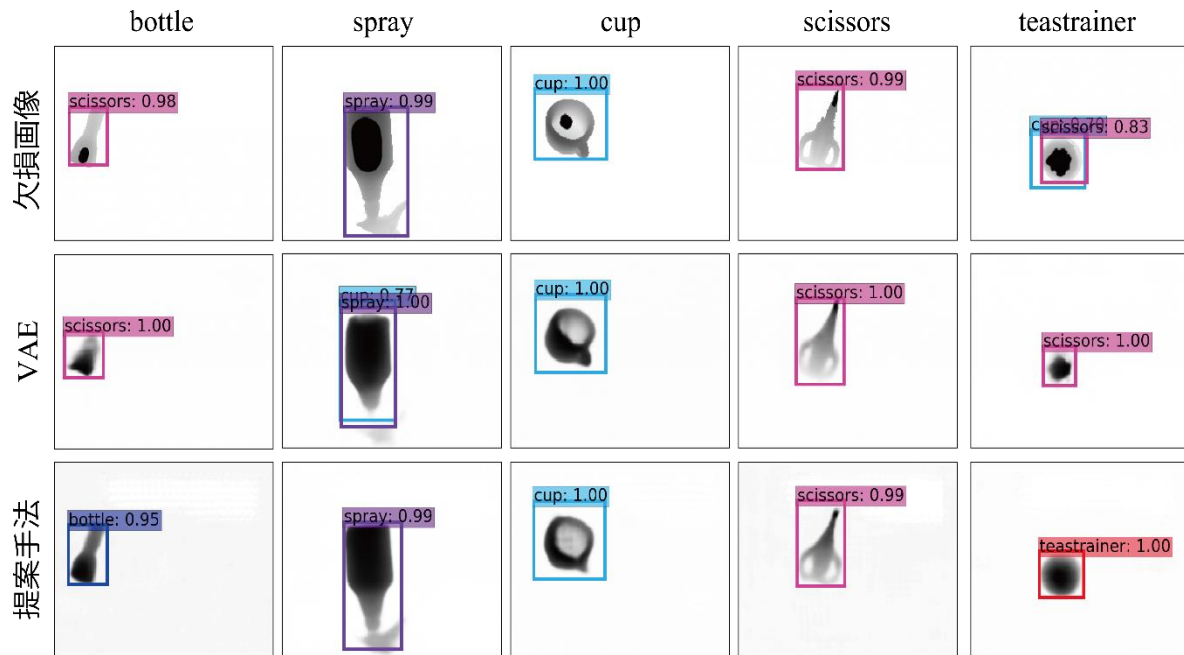


図 7. 検出結果の例.

Autoencoders for Photographic Image Synthesis, NIPS, pp.52-63, 2018.

[13] Nathan Koenig, Andrew Howard : Design and Use Paradigms for Gazebo, An Open-Source Multi-Robot Simulator, IROS, 2004

[14] ORGE-Open Source 3D Graphics Engine. 「Home」
<https://www.ogre3d.org/>

[15] Wei Liu et al. : SSD: Single Shot MultiBox Detector, ECCV, 2016.

本田圭：中部大学工学部ロボット理工学科在学。現在，距離画像の欠損を補間する研究を進めている。

三輪顕太郎：中部大学工学研究科ロボット理工学専攻在学。現在，距離画像の欠損を補間する研究を進めている。

山内悠嗣：2012 年中部大学大学院博士後期課程修了。2010 年-2012 年 独立行政法人日本学術振興会特別研究員 DC2。2014 年中部大学助手，2017 年より同大学講師。画像認識，機械学習，知能ロボティクスの研究に従事。