一般社団法人　電子情報通信学会
THE INSTITUTE OF ELECTRONICS,
INFORMATION AND COMMUNICATION ENGINEERS

信学技報
IEICE Technical Report
PRMU2017-87(2017-10)

姿勢を表現する多様体に基づくGANsを用いた物体姿勢推定の検討

川西　康友† 　出口　大輔†† 　井手　一郎† 　村瀬　　洋†

† 名古屋大学 大学院 情報学研究科　〒464-8601 愛知県名古屋市千種区不老町
†† 名古屋大学 情報連携統括本部 情報戦略室　〒464-8601 愛知県名古屋市千種区不老町
E-mail: †kawanishi@i.nagoya-u.ac.jp

あらまし　Generative Adversarial Nets（GANs）[1] は，ある事前分布から画像などのデータを生成可能なネットワークであり，乱数から様々なデータを生成可能である．本研究では，物体の姿勢変化は潜在的には多様体で表現可能であることに着目し，多様体上に定義した確率分布から画像を生成するGANsの枠組みと，それを用いて学習画像を補間しつつ姿勢推定器を学習する手法を提案する．実験では，全周を撮影した複数の物体に対し，提案手法により学習データに含まれない姿勢が補間できることを確認した．また，姿勢推定を同時に学習し，学習データに含まれない姿勢に対する姿勢推定の可能性を評価し，学習画像の枚数が少ない場合でも姿勢推定精度が大幅には低下しないことを確認した．

キーワード　多様体，敵対的生成ネットワーク，パラメトリック固有空間法

A Study on GANs based on Pose Manifold for Rigid Object Pose Estimation

Yasutomo KAWANISHI†, Daisuke DEGUCHI††, Ichiro IDE†, and Hiroshi MURASE†

† Graduate School of Informatics, Nagoya University
Furo-cho, Chikusa-ku, Nagoya, Aichi, 464-8601, Japan
†† Information Strategy Office, Nagoya University
Furo-cho, Chikusa-ku, Nagoya, Aichi, 464-8601, Japan
E-mail: †kawanishi@i.nagoya-u.ac.jp

Abstract　Generative Adversarial Nets (GANs) is a pair of neural networks which can learn data distribution and generate various data from the distribution. In this research, by focusing on the fact that pose variation of a rigid object can be expressed on a manifold in a latent space, we introduce a GANs model which generates data from a distribution defined over a manifold. We also propose a pose estimation method which trains a pose estimator while interpolating training images using the GANs. We evaluated the interpolation capability of the proposed model using a public dataset, and also evaluated pose estimation accuracy of the proposed model.

Key words　Manifold, Generative Adversarial Nets, parametric eigenspace

1. はじめに

産業や生活支援など様々な分野でロボットが導入されつつある．産業分野では，ロボットの導入による生産現場での部品のピッキングに関する自動化の需要が高まっており，自動ピッキングロボットの性能を競う，Amazon Picking Challange [2] と呼ばれるコンペティションが開催されている．また，生活支援においても，少子高齢化の進展に伴う介護・福祉，家事等の労働力不足の懸念から，日常的な生活支援を目的としたロボットが開発されている [3]．生活支援におけるロボットの主な役割は，人の指示に従って，物体を受け渡しすることである．これらの分野とも，ロボットが物体を把持することが共通の課題であり，物体をうまく把持するための技術が求められている．ロボットが物体をうまく把持するには，対象物体をうまく把持できる方向を知るため，物体の姿勢推定を行なう必要がある．

物体の姿勢推定を実現するためにロボットに搭載するセンサとしては，Kinect や Xtion などに代表される 3 次元ビジョンセンサなどが考えられる．しかし，価格が高い，サイズが大きい，使用環境やデータ取得可能な対象物体が限定的，キャリブレーションが煩雑など，実用上の課題も多く，3 次元ビジョンセンサではなく従来の単眼カメラからの物体の姿勢推定技術が求められている．

一方，深層学習を用い，乱数から様々なデータを生成でき

This article is a technical report without peer review, and its polished and/or extended version may be published elsewhere.
Copyright ©2017 by IEICE

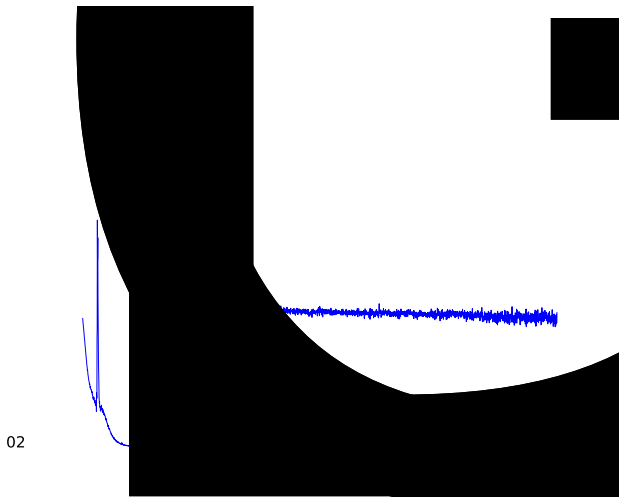

図 8　学習の過程.

図 6　姿勢の教師信号を与えて学習した GANs により生成したもの.

図 7　データセットから選択した画像.

表 4　姿勢推定結果

	10 度刻み	20 度刻み	40 度刻み
提案手法	11.33 度	12.77 度	27.19 度
比較手法	9.77 度	11.57 度	32.97 度

4.4　姿勢推定実験

次に，3.3 節で提案した，姿勢推定器を同時に学習する GANs について，学習データの数に応じた姿勢推定精度の変化を調べた.

4.4.1　実 験 方 法

COIL-20 データセットに含まれる各物体に対し，0 度の角度から 10 度刻み，20 度刻み，40 度刻みに画像を学習用画像として選択し，5 度の角度から 10 度刻みに選択した画像を評価に用いて姿勢推定の精度を平均絶対誤差で評価した.

比較手法として，GANs による画像生成をせず，提案手法で用いた Discriminator の姿勢推定部分からなる CNN を用いた．このネットワークは，入力となる対象物体の実画像のみを用いて学習した.

4.4.2　結　　果

まず，10 度刻みで学習した時の，学習の様子を図 8 に示す．判例中の，Classification loss 及び Orientation loss は，実画像か生成画像かの判定及び，物体の姿勢推定の，Discriminator と Generator のそれぞれにおける誤差を表す．生成した画像が本物と見分けがつかなくなり，姿勢推定誤差も小さくなることが分かる.

次に，姿勢推定結果を表 4 にまとめる.

提案手法が比較手法である通常の CNN による推定よりも精度が低いのは，4.3.2 節で述べたように，生成した画像が本来の向きから少しずれることがあるため，提案手法は誤った姿勢を学習したためと考えられる．一方，学習画像の数が少なくなっても，比較手法に比べて姿勢推定精度の低下が小さいのは，GANs によって画像を補間することにより，学習画像の不足を補いながら推定器を学習出来ているからだと考えられる.

5.　お わ り に

本報告では，物体の姿勢変化は潜在的には多様体で表現可能であることに着目し，多様体上に定義した確率分布から画像を生成する GANs の枠組みと，それを用いて学習画像を補間しつつ姿勢推定器を学習する手法について検討を行なった.

実験では，教師信号を与えて学習することにより，物体の連続的な姿勢変化に応じた画像を生成できることを確認したが，真値からずれた画像が生成されてしまうことがあった．これは，画像の生成と姿勢推定を同時に学習させたことにより，まだうまく生成出来ていないデータを用いて姿勢推定を学習したことが原因の 1 つであると考えられる．今後の課題として，学習の方法及び，教師信号の与え方について検討する必要がある.

本報告では特定物体の姿勢推定について扱ったが，今後の課題として，例えば姿勢を表現する多様体を 2 次元平面中の円周から 3 次元以上の空間中の円環に変更することにより，同一クラスに属する様々な物体の姿勢変化を扱えるような拡張が考えられる.

謝辞　本研究の一部は，科学研究費補助金による.

文　　献

[1] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, “Generative adversarial nets,” Advances in Neural Information Processing Systems 27, pp.2672–2680, Dec. 2014.

[2] N. Correll, K.E. Bekris, D. Berenson, O. Brock, A. Causo, K. Hauser, K. Okada, A. Rodriguez, J.M. Romano, and P.R. Wurman, “Lessons from the Amazon picking challenge,” arXiv preprint arXiv:1601.05484, pp.1–14, Jan. 2016.

[3] J. Broekens, M. Heerink, and H. Rosendal, “Assistive social robots in elderly care: A review,” Gerontechnology, vol.8, no.2, pp.94–103, April 2009.

[4] H. Murase and S.K. Nayar, “Visual learning and recognition of 3-D objects from appearance,” Int. J. of Comput. Vision, vol.14, no.1, pp.5–24, Jan. 1995.
[5] S. Gupta, P. Arbelaez, R. Girshick, and J. Malik, “Aligning 3D models to RGB-D images of cluttered scenes,” Proc. IEEE Conf. on Comput. Vision and Pattern Recognit., pp.4731–4740, June 2015.
[6] M. Torki and A. Elgammal, “Regression from local features for viewpoint and pose estimation,” Proc. 2011 Int. Conf. on Comput. Vision, pp.2603–2610, Nov. 2011.
[7] **二宮宏史, 川西康友, 出口大輔, 井手一郎, 村瀬　洋, 小堀訓成, 橋本国松, “深層学習を用いた多様体構築による 3 次元物体の姿勢推定に関する予備検討,” 信学技報, 第** 116 **巻,** pp.25–30**,** Jan. 2016**.**
[8] A. Krizhevsky, I. Sutskever, and G.E. Hinton, “Imagenet classification with deep convolutional neural networks,” Advances in Neural Information Processing Systems 25, pp.1097–1105, Dec. 2012.
[9] M. Cimpoi, S. Maji, I. Kokkinos, and A. Vedaldi, “Deep filter banks for texture recognition, description, and segmentation,” Int. J. of Comput. Vision, vol.118, no.1, pp.65–94, May 2016.
[10] A. Radford, L. Metz, and S. Chintala, “Unsupervised representation learning with deep convolutional generative adversarial networks,” arXiv preprint arXiv:1511.06434, pp.1–16, Nov. 2015.
[11] M. Mirza and S. Osindero, “Conditional generative adversarial nets,” arXiv preprint arXiv:1411.1784, pp.1–7, Nov. 2014.
[12] P. Isola, J.-Y. Zhu, T. Zhou, and A.A. Efros, “Image-to-image translation with conditional adversarial networks,” arXiv preprint arXiv:1611.07004, pp.1–16, Nov. 2016.
[13] A. Shrivastava, T. Pfister, O. Tuzel, J. Susskind, W. Wang, and R. Webb, “Learning from simulated and unsupervised images through adversarial training,” Proc. IEEE Conf. on Comput. Vision and Pattern Recognit., pp.2107–2116, July 2017.
[14] C. Wan, T. Probst, L. Van Gool, and A. Yao, “Crossing nets: Combining GANs and VAEs with a shared latent space for hand pose estimation,” Proc. IEEE Conf. on Comput. Vision and Pattern Recognit., pp.680–689, July 2017.
[15] S.A. Nene, S.K. Nayar, and H. Murase, “Columbia object image library (COIL-20),” Technical Report CUCS-005-96, Department of Computer Science, Columbia University, Feb. 1996.