

DCGAN を用いた記憶と表象のモデル

Deductive model for memory and representation using DCGAN

小島大樹 *1 土井樹 *1 シナパヤ ラナ *1 池上高志 *1
 Hiroki Kojima Itsuki Doi Lana Sinapayen Takashi Ikegami

*1 東京大学総合文化研究科

Graduate School of Arts and Sciences, University of Tokyo

There is a growing trend to understanding perception as a generative process, which actively construct a internal model of the world. Some studies proposed artificial models to capture this characteristics of perception, but implementing such models have been quite difficult due to the lack of suitable artificial generative systems. Here, we propose a new model of perception using powerful generative system, DCGAN. DCGAN is one of the deep neural network system, and can generate new images from input image dataset. We show that this model might be able to construct a model of the world by providing local view images and discuss new ways to study perception using this kind of deductive models.

1. はじめに

知覚を、感覚入力のフィルターのように受動的に捉えるのではなく、内部モデルを生成しそれをベースとして外界を捉える。こうした見方が、特に Helmholtz(1867)以降の、知覚を推論として扱う立場において現れてきた。

内部モデルがあるかないかは、物理的実体としての生物の脳を調べることは不可能である。そこで人工的に内部モデルを生成して、ナビゲーションを行なうシステムを作って解析するというアプローチが考えられる。例えば Rössler(1981)[Rössler 1981]ではこうした考えに基づき、人工的なナビゲーションの青写真を提案している。この Rössler の提案時点においては、十分な内部モデルの生成能力をもつシステムがなかったために、実際にそのようなナビゲーションシステムを構築し、研究することは不可能であった。近年になって、Deep Neural Network(DNN)の発展の中で、画像認識の能力が飛躍的に向上し、それに基づいた、生成系の提案が行われるようになってきた。特に deep convolutional generative adversarial network (DCGAN) と呼ばれるモデル・システムは、イメージの生成系と、イメージの判定系を共進化させることで、高精度のイメージ画像を生成することを可能にしている。

そこで今回、この強力な生成系をベースに情報の入力列から外界の内部イメージを生成し、それをもとにナビゲーションを行い、というフィードバック・モデルを考える。次節以降で、DCGAN に関する簡単な解説と、それによるわれわれの実験における DCGAN の生成イメージを報告する。その上で可能な知覚モデルを提案し、議論する。

2. Model

2.1 ニューラルネットワーク: DCGAN

Discriminator(D) と Generator(G) を共進化させることで、元の入力データの分布を知ることなく、同様の分布をもとにイメージを出力するシステムを GAN[Goodfellow 2014]、特に深層の畳み込みニューラルネットワーク (Deep Convolution Neural Network) を使ったシステムを DCGAN[Radford 2015] という。GAN の目標は、学習データセットと見分けがつかない

いデータを生成する Generator (G) を獲得することである。G には、一様分布などからサンプルされた乱数 z が入力され、これを種として画像を生成する。G の学習には、Discriminator (D) を使われる。D は入力が学習データセット由来か、G の生成した自動生成データを判別する classifier となっている。DCGAN では、D は通常の Convolutional neural network、G は z からスタートする逆方向の CNN を用いる。

本研究では Radford らのモデルを元に、より大きな画像を扱えるようにするために入力と出力の画像データを 64×64 から 96×96 に変更し、計算コストを抑えるためにいくつかの変更を行ったものを使用した。

2.2 画像空間 (z space)

DCGAN の学習においては、どの z の値から生成した画像も、D によって本物と判定されるようなものになっていくと同時に、出力画像は z に対して連続的に変化していくため、学習の結果として、 z をパラメタとして連続的に変化する画像の空間ができあがる。この空間は、DCGAN の構造に加え、データセットそのものの持つ構造によるものであると考えられる。

本研究ではこの z 空間が、内部イメージに相当するものであると考え、その観点から z 空間に出来上がる構造の解析を行うこととした。

また、通常 z 空間を 100 次元にとることが多いが、ここでは z 空間の解析を簡単にするため、 z 空間の次元を 5 次元に設定して学習を行った。

2.3 データセット

z 空間は学習データセットの性質に強く依存すると考えられるため、どのようなデータセットを使うかということが重要である。通常の DCGAN では一定の基準を設けて、その範囲内で無差別に画像を収集したものを学習データセットとして用いることが一般的である。しかし、通常の状況で人が受け取る視覚情報は、一人称的な視点に基づくものである。

そこで本研究では、このデータセットの性質の違いが本質的なものであると考え、一人称的な視点に基づく画像のデータセットを使用して DCGAN の学習を行い、これにより自然に外界の内部モデルが z 空間上に構成されるのではないかと考えて実験を行った。

このデータセットとしてここでは、一人称視点の動画を撮影

し、このフレームを画像として取り出したものを用いることとした。

3. 結果

3.1 DCGAN の出力

一人称的な画像のデータセットを DCGAN に学習させることにより、以下の様な出力を得た。

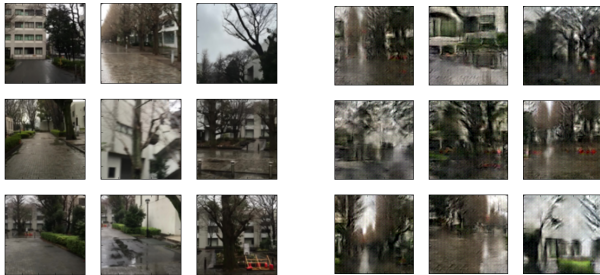


図 1: input 画像の例 (左) と output 画像の例 (右)

3.2 z 空間内での実空間構造

また z 空間において、z の値を連続的に変化させていくことで、実空間を少しずつ移動していく時に見える風景に対応するものが得られることがわかった。



図 2: z の値を直線的に変化させた時に得られる出力画像の例

上の図では、z の値を [-1.0, 0.2, -1.0, -1.0, -1.0] から、[1.0, 0.2, 1.0, 1.0, -1.0] に直線的に変化させている。

4. 議論

4.1 ナビゲーションへの拡張

今回の結果より、一人称視点の画像のデータセットの学習を行うだけで、外界についての内部モデルが構築される可能性が示唆された。しかし、実際にナビゲーションとして機能するためには、まず、現在見ている風景から、内部モデル内での位置への変換を行う必要がある。今回のモデルでは、画像から対応する z 空間上の位置への変換は組みこまれていないが、この変換もあわせて学習するモデルとして VAEGAN[Larsen 2015]がある。本発表においては、このモデルを用いて学習させた場合の結果についても議論する予定である。

また、ナビゲーションのためには、ゴール地点を内部モデル上で表現し、現在の位置から、その地点まで導いていくことが必要である。今回のモデルには、体の動きに対応するものが直接入っていないため、実際の運動と、z 空間内の軌跡の対応付けが必要である。いくつかの方法が考えられるが、例えば、(生体における grid cell のような) 純粋に運動にもとづく map との関連付けをおこなうか、vicarious trial and error のように各地点での首振り運動と、それに伴う画像の変化を見ることが、逐次的に z 空間と運動の間の対応付けをおこなうといった方法が考えられる。

4.2 place cell 研究との関係

cognitive map の研究としては、海馬内の構造である place cell と関連付けて論じられることが多く [Tolman 1948] [O'keefe 1978]、現在も活発に研究がすすめられている。ただし、こういった研究は Morris の water maze のような、単純な目印を持つ小さい空間内で行われることが多いため、実世界のように視覚刺激が複雑な場合や、さらに広い空間でのナビゲーションについては不明瞭な点が残っており、これらの点について、今回のモデルのようなアプローチが有効な可能性があると考えている。

また、place cell のモデルにおいて、単純な visual cue ではなく実際の画像を用いた研究として、例えば [Strösslín 2005] があり、これらの研究との関連についても議論する予定である。

参考文献

- [Rössler 1981] Rössler, Otto E. "An artificial cognitive map system." *BioSystems* 13.3 (1981): 203-209.
- [Goodfellow 2014] Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., ... and Bengio, Y. (2014). Generative adversarial nets. In *Advances in Neural Information Processing Systems* (pp. 2672-2680).
- [Radford 2015] Radford, A., Metz, L., and Chintala, S. (2015). Unsupervised Representation Learning with Deep Convolutional Generative Adversarial Networks. arXiv preprint arXiv:1511.06434.
- [Larsen 2015] Larsen, A. B. L., Sønderby, S. K., and Winther, O. (2015). Autoencoding beyond pixels using a learned similarity metric. arXiv preprint arXiv:1512.09300.
- [Tolman 1948] Tolman, Edward C. "Cognitive maps in rats and men." *Psychological review* 55.4 (1948): 189.
- [O'keefe 1978] O'keefe, John, and Lynn Nadel. *The hippocampus as a cognitive map*. Vol. 3. Oxford: Clarendon Press, 1978.
- [Strösslín 2005] Strösslín, Thomas, et al. "Robust self-localisation and navigation based on hippocampal place cells." *Neural networks* 18.9 (2005): 1125-1140.