

深層学習とパーシステントホモロジーによるハイブリッド画像解析

鍛冶 静雄¹

¹九州大学マス・フォア・インダストリ研究所

e-mail : skaji@imi.kyushu-u.ac.jp

畳み込みニューラルネットワーク (CNN) は画像解析において非常に強力な道具である一方で、近視眼的でありテクスチャーなど局所的な情報に捉われるすぎる傾向があることが知られている。これは、人間がものを見る時にまず大域的な手がかりを重視するのと対照的である。本講演では、パーシステントホモロジー (PH) によって画像の大域構造を、CNN が学習しやすい形にエンコードする手法を紹介する。2通りのものの見方を融合することで、実際に画像認識精度が向上することを例示する。

現在の AI ブームは画像認識で深層学習が大きな性能を示したことに始まった。画像に用いられる深層学習の基本は畳み込み演算であるが、これは画像を2変数関数とみなし、小さな台を持つ関数 h を畳み込む演算である。この h に依存して画像の様々な局所的特徴が抽出される。つまり、深層学習では主に画像の局所的特徴に着目していると言える。実際、深層学習では工夫を加えなければ形状といった大域的な情報よりも、テクスチャーなど局所的情報を優先して学習することが示されている [1]。一方、人間が画像を認識する際には、位相的な不変量で捉えることができるような、大まかな様子をまず手がかりとするであろう。とはいえ、直感に反して位相はノイズに対して頑強ではない。二値画像を考えると、たった1ピクセルの差で位相が変わる例がいくつでも作れる。そこで、画像の位相的な解析では、頑強な不変量として位相データ解析の主要な道具である PH が使われる。

深層学習は完全にデータドリブンで圧倒的な性能を実現するが、いくつかの問題を抱えている。特に重大なものを二つあげると：

- 大量の学習データが必要である。(医療や土木など分野によってはデータ取得コストが高く、いわゆるビッグデータを用意できない)
- 説明性と安定性に欠ける。(人の目にはわからないほどの微小なノイズを加えることで結果を変えてしまう、adversarial

example [2] などの深刻な攻撃が知られている)

これらの問題を軽減するには、いわゆるルールベース (演繹的) のトポロジーを基礎とする方法との融合が有効であろう。その一つの実現法を紹介するのがこの講演の目的である。

パーシステントホモロジーは、位相空間の上の実数値関数に対して定義される不変量で、関数へのノイズに対する頑強性を持つ。一般的な定義は [3] など良いサーベイが豊富にあるのでそちらに譲るとして、ここでは画像に特化した場合を紹介する。ここで画像とは、2次元のモノクロ階調画像もしくは、CT スキャンなど3次元のボリュームデータを指すこととする。この設定でサイズ N の画像は、有限の正方格子 $\Omega = \{(x, y, z) \in \mathbb{Z}^3 \mid 0 \leq x, y, z \leq N - 1\}$ (の部分格子) 上で定義された実数値関数 $f : \Omega \rightarrow \mathbb{R}$ とみなせる。二次元画像の場合は f の定義域は $z = 0$ なる部分格子に含まれるものとする。 Ω は *cubical complex* の構造を持つ: 0セルは頂点、1セルは辺、2セルは正方形、3セルは立方体であり、自明な方法で境界写像が定義される。実数 $\alpha \in \mathbb{R}$ を指定するごとに、 Ω の部分複体 $f^{-1}((-\infty, \alpha])$ が定まる。セル σ は、それを構成する全ての頂点について f の値が α 以下であるとき、 $f^{-1}((-\infty, \alpha])$ に属するという定義である。 f の値域は有限個の実数からなる集合であるが、これを小さい順に $\alpha_1, \dots, \alpha_m$ とすると、ホモロジーの列

$$\begin{aligned} H_i(f^{-1}((-\infty, \alpha_1])) &\rightarrow H_i(f^{-1}((-\infty, \alpha_2])) \rightarrow \dots \\ &\rightarrow H_i(f^{-1}((-\infty, \alpha_m])) \rightarrow H_i(f^{-1}((-\infty, \alpha_{m+1}))) \end{aligned}$$

が得られる。ここで便宜上、 $\alpha_{m+1} = \infty$ とした。ホモロジーの係数を体にとっておくと (実用上は \mathbb{R}_2 が選択されることが多い)、これは一意的に既約分解¹することができ、いわゆるバーコード $[\alpha_i, \alpha_j) \in \mathbb{R} \times (\mathbb{R} \cup \infty)$ の集合で記述される。端的には、PH は画像を入力として区間の集合を出力する写像・特徴抽出機とみなせる。画像の PH を計算するソフトウェアはいくつか存在

¹係数体上1変数多項式環の上の加群として

するが、講演者も開発に携わっている **Cubical Ripser** [4] が現時点で最速であると思われる。オープンソースで公開されており、Python から手軽に利用できるのも興味のある方は是非試していただきたい。Cubical Ripser の利用を含めた位相データ解析のチュートリアルも用意している [5]。

他方で、深層学習の基本構成要素である畳み込みは、複数枚の同一サイズの画像を入力にとり、複数枚の同一サイズの画像を出力する線型変換である。入力と出力で枚数やサイズは一致しているとは限らない。複数枚のそれぞれはチャンネルと呼ばれ、例えば一般的なカラー画像であれば、赤緑青を表す3チャンネルで一枚の画像を構成する。畳み込みの出力の各点の値は、入力画像内の小さな矩形領域にのみ依存するという意味で、局所的な演算である。深層学習で用いられる CNN は、畳み込みとピクセル毎の単純な非線形関数の何重もの合成写像であるが、畳み込み写像のパラメーターが学習によって決定される。詳しい定義と具体例は、[6, §1] にまとめた画像に対する深層学習の基本を参照されたい。

PH の大域性と深層学習の局所性という双方の強みを活かすことを目指し、それらを併用した画像解析手法としては [7] などが知られているが、特別な CNN を用意する必要があること、PH の持つ座標情報を利用しないことなど、改善の余地がある。ここでは、PH で抽出した情報を画像の形にエンコードする手法を紹介する。得られた画像は元の画像にチャンネルとして追加することで、通常の CNN でそのまま処理できる。画像の大域情報を PH で抽出し、畳み込みで捉えられる形でアノテーション化するというのが基本アイデアである。PH の情報を保持する画像は次の様にして得られる。各 $k = 0, 1, 2$ に対して k 次元 PH はバーコード $[\alpha_i, \alpha_j]$ の集合を与えるが、さらに一つ一つのバーコードにはそれがどのセルによって生まれたか (もしくは死んだか) を対応づけることができ、つまり画像中の座標の情報を持つ。そこで、各 k 毎に次の様な2チャンネルの画像を考えることができる: 各ピクセルに対して、その座標を持つバーコード $[\alpha_i, \alpha_j]$ が存在すれば²、 α_i, α_j をそのピクセル値とする (図1参照)。

この手法は、土木と医療における実問題に

²複数存在する場合は、寿命 $\alpha_j - \alpha_i$ の大きいものを優先する

用されて既に成果を上げているが、諸事情あってその詳細はまだ書くことはできない。代わりにシンプルな適用例として手書き数字の識別問題を紹介する。0 から 9 の 10 種の種類に対して、わずか1例ずつの教師画像により CNN を学習させる実験を行った [8]。このように極端に学習データが少ない場合に、PH を CNN に援用することで精度の向上 (52% → 55%) が見られることが確かめられた。

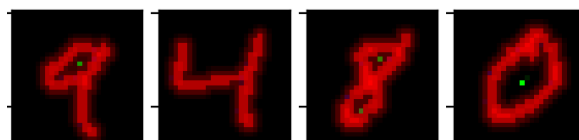


図1. 1次元パーシステントホモロジーの情報(寿命)を緑チャンネルとして、赤チャンネルの数字画像と重ねた画像。大域的なトポロジー情報が局所的なピクセルの色としてエンコードされている。

参考文献

- [1] R. Geirhos et al., *ImageNet-trained CNNs are biased towards texture; increasing shape bias improves accuracy and robustness*, ICLR 2019.
- [2] C. Szegedy et al., *Intriguing properties of neural networks*, ICLR 2014.
- [3] H. Edelsbrunner and J. Harer, *Computational Topology - An Introduction*, American Mathematical Society, 2010.
- [4] S. Kaji, T. Sudo, K. Ahara, *Cubical Ripser: Software for computing persistent homology of image and volume data*, arXiv:2005.12692, codes available at https://github.com/shizuo-kaji/CubicalRipser_3dim.
- [5] *Tutorial on Topological Data Analysis*, <https://github.com/shizuo-kaji/TutorialTopologicalDataAnalysis>.
- [6] S. Kaji and S. Kida, *Overview of image-to-image translation by use of deep neural networks: denoising, super-resolution, modality conversion, and reconstruction in medical imaging*, Rad. Phys. and Tech., Vol. 12(3) (2019), 235–248.
- [7] C. Hofer, R. Kwitt, M. Niethammer, and A. Uhl, *Deep Learning with Topological Signatures*, NIPS2017, 2017.
- [8] *Homology-assisted Convolutional Neural Networks for 2D/3D image classification*, <https://github.com/shizuo-kaji/HomologyCNN>.