The 33rd Annual Conference of the Japanese Society for Artificial Intelligence, 2019

画像の “写真らしさ” に関する数学的アプローチについて

On a mathematical approach to “photo-likeness” of images

浅尾　泰彦 *1　　坂本　龍太郎 *1
Yasuhiko Asao　　Ryotaro Sakamoto

*1東京大学大学院数理科学研究科
Graduate School of Mathematical Science, the University of Tokyo

In image recognition, it is significant to determine the boundary between meaningful and non-meaning images. In this paper, we show a mathematical approach to this problem by defining a “quasi-photographic” image. In order to formulate the question ‘What is photograph likeliness ? ’ mathematically, we introduce a function ‘depth’ that takes real values for images and analyze its asymptotic behavior. We also examine that an actual photograph is indeed a quasi-photograph. The idea of depth comes from the rank of the 0th persistent homology of a cubical complex and it can be expected that more precise classification of images can be obtained by analyzing the higher rank in the future. We also believe that it can be applied to deep learning, which is being actively utilized recently in image recognition, to selection of learning data. We would like to propose one approach of applicating pure mathematics in image recognition.

1.　はじめに

本稿で扱う問題は，コンピュータが “内在的” にどこまで意味のある画像とそうでないものを区別できるか？ というものである．

以下では意味のある画像のことを「写真」と呼ぶことにする．つまり絵や数字など，我々が目にしてすぐに意味が理解できる，もしくは意味があると判断できるもの全てを「写真」と総称する．一方で「意味のない」画像というのは一見して意味が理解できない，意味があると判断できないものを指す．例えば図 1 右のようなモザイク画像のことを指している．

機械学習ではコンピュータに性質 A を持つ大量の類似データを学習させることで，新たに読み込ませたデータが性質 A を持つかどうかを判断させることができた．例えばりんごの写真を学習させた後に図 1 右の画像を読み込ませると，コンピュータはそれがりんごでないと判断できる．我々がここで「内在的」と言っているのは，そのような学習の過程を経ないで，ということである．つまりりんごが何であるかを知らない状況で，コンピュータはりんごの写真とモザイクの画像をきちんと分類することが可能であるか？ また可能であれば「りんごとモザイク」という極端な分類の他にどの程度分類が可能であるか？

本稿では，画像の持つ数理的な性質によって特徴付けられる「準写真」という画像のクラスを導入することで，この問題に取り組んだ．準写真であるという性質は個々の画像に対して数学的に有無が判別できるため上で述べた意味で内在的であり，従ってコンピュータは学習の過程を経ずに画像を準写真とそうでないものに分類することができる．

さらに数学的に定義された準写真であるという性質は，実際の写真にもきちんと備わっていることを例で確かめることができた．

準写真は画像の「深さ」という数学的概念を定義することで得られ，深さは近年データサイエンスの分野で広く認知されているパーシステントホモロジーから着想を得ている．本稿においてパーシステントホモロジーなど純粋数学で成熟した道具を，画像認識に活用する 1 つのアプローチを提案したい．

図 1: 右は depth が非常に大きい．

2.　画像の深さ

2.1　画像の定式化

$[0,1)$ で 0 以上 1 未満の実数の集合を表す．自然数 N に対して集合 $\square_N$ を

$$\square_N := \left\{ \frac{0}{2^N}, \frac{1}{2^N}, \dots, \frac{2^N-1}{2^N} \right\}^2 \subseteq [0,1) \times [0,1).$$

で定義する．$C = \{0, 1, \dots, n-1\}$ を白黒の濃淡を表す集合とする．$2^N \times 2^N$ ピクセルのモノクロ画像は，写像 $\square_N \to C$ そのものである．ここで $2^N \times 2^N$ ピクセル画像を図 2 のように $2^d \times 2^d$ 分割することを考える．すなわち $\square_N$ の部分集合族 $\square_N^d$ を

$$\square_N^d := \left\{ \frac{1}{2^d}\square_{N-d} + x \subseteq \square_N \middle| x \in \square_d \right\}.$$

で定め非交叉和による $2^d \times 2^d$ 分割 $\square_N = \bigsqcup_{\square \in \square_N^d} \square$ を与える．

例えば，$\square_N^0 = \{\square_N\}$，$\square_N^N = \{\{x\} \mid x \in \square_N\}$ である．

連絡先: 浅尾泰彦： asao@ms.u-tokyo.ac.jp ，
坂本龍太郎： sakamoto@ms.u-tokyo.ac.jp

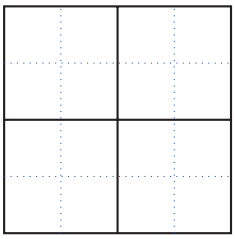

図 2: 実線は $\square_2^1$, 点線は $\square_2$.

図 3: $N = 10$ のキリンの画像．左上から右下にかけて $0 \leq d \leq 10$ が大きくなっていく．$d = 5$ で初めて黒いマスが現れるので depth $= 0.5$ である．

2.2 深さの定義

2 色モノクロ画像 $f : \square_N \to C = \{0, 1\}$ に対して，どれくらい色の偏りがあるかを測る指標として次の関数 φ_d を定義する．

$$\varphi_d(f) := \min_{\square \in \square_N^d} \left(\sum_{i,j \in \square} |f(i) - f(j)| \right).$$

つまり φ_d は $2^d \times 2^d$ 分割したそれぞれのマスのうち，もっとも色が偏ったマスの偏り具合を数値として表す関数である．1 つでも「全て白（黒）」であるマスが存在すればその値は 0 となる．特に $\varphi_N = 0$ が常に成り立つ．

（2 色と限らない）画像 f の深さを，どれくらい分割を細かくした時に（d を N に近づけたときに）初めて $\varphi_d = 0$ となるか，を測ることで定義する．

定義 1 $2^N \times 2^N$ ピクセル n 色画像 $f : \square_N \to C$ に対して

$$\mathrm{depth}(f) := \frac{1}{N} \min\Big\{ d \in \{0, \ldots, N\} \Big| 0 = \min_{p : \mathrm{Im} f \to \{0,1\}} \varphi_d(p \circ f) \Big\},$$

を f の**深さ**という．ここで p は f の像から集合 $\{0, 1\}$ への全射全体を動く．

例えば $\mathrm{depth}(f)$ が 0 であることと，f が定値写像であること（つまり 1 色画像）は同値である．逆に図 1 右のような複雑な画像の depth は 1 となる．つまり depth は画像の複雑さを測る指標となっている．意味のある画像はある程度色に偏りがあると考えられるため depth は低いことが期待される．図 3 は d を次第に大きくした時の $\left(\sum_{i,j \in \square} |f(i) - f(j)| \right)_{\square \in \square_N^d}$ を色の濃さとして表したものである．d が増加するとマスは細かくなり，初めて真っ黒なマスが現れた時の d が depth に対応する．

3. 深さの漸近挙動と準写真の定義

この章では前章で定義した depth を用いて，どれくらいの割合の画像が「とても複雑」（つまり写真らしくない）かを計算する．それに基づいて準写真という数学的に「写真らしい」画像のクラスを定義する．

$\mathrm{Map}(\square_N, C)$ で $\square_N$ から C への写像全体（つまり $2^N \times 2^N$ ピクセルモノクロ画像全体）の集合を表す．$0 \leq \alpha \leq 1$ に対して depth が α 以下の画像全体の集合を $\mathcal{P}_{N,C}(\alpha)$ とおく．有限集合 X の要素の個数を $\#X$ で表すことにすれば，$\#\mathrm{Map}(\square_N, C) = n^{4^N}$ より，depth が α 以下の画像全体の割合は $\frac{\#\mathcal{P}_{N,C}(\alpha)}{n^{4^N}}$ である．このとき次の計算結果を得る．

命題 1

$$\lim_{N \to +\infty} \frac{\#\mathcal{P}_{N,C}(\alpha)}{n^{4^N}} = \begin{cases} 0 & \text{if } \alpha < 1, \\ 1 & \text{if } \alpha = 1. \end{cases}$$

すなわち「ほとんど全ての画像がとても複雑」であることがわかる．この結果は，ランダムに生成した画像が写真であることは稀有である，という我々の認識に関する直観と矛盾しない．

上の結果を $\#\mathcal{P}_{N,C}(\alpha)$ の主要項の漸近挙動だと考え，第 2 主要項を計算すると以下の結果を得る．

命題 2

$$\lim_{N \to +\infty} \frac{\#\mathcal{P}_{N,C}(1 - \alpha \frac{\log N}{N})}{n^{4^N}} = \begin{cases} 0 & \text{if } 1/\log 4 < \alpha, \\ 1 & \text{if } 0 \leq \alpha \leq 1/\log 4. \end{cases}$$

ただし log の底は 10 とする．

従って画像全体のうち depth が $1 - \frac{\log N}{N \log 4}$ 未満であるような「複雑すぎないもの」は非常に少ないことがわかる．前章でも述べたように意味のある画像は depth が低いと期待されるため，これらを準写真と定義する．

定義 2 $2^N \times 2^N$ ピクセル n 色画像 $f : \square_N \to C$ が

$$\mathrm{depth}(f) < 1 - \frac{\log N}{N \log 4}$$

を満たすとき，f は**準写真**であるという．

$1 - \frac{\log 10}{10 \log 4} = 0.93979400086\cdots$ より，図 3 のキリンの画像（depth $= \frac{5}{10} = 0.5$）は準写真である．

4. 人間の認識に関する予想

図 3 では左上の真っ白な画像から始まって次第に細度が上がっていきキリンの姿が浮かび上がり，最後には真っ黒な画像になる．真っ黒になる 1 つ手前の画像はもうほぼ写真と変わりないが，もう 1 つ手前までいくと写真と認識できるもののやや画質が悪いという印象を持つ．これら 11 枚の画像に真っ白なものから順番に $\frac{0}{10}, \frac{1}{10}, \ldots, \frac{10}{10}$ と数を振ると，$\frac{8}{10}$ と $\frac{9}{10}$ の間がちょうど画質の良し悪しを判断する境目ということになる．我々の予想はこの境目の値がおおよそ $1 - \frac{\log N}{N \log 4}$ に対応するのではないかというものである．実際，前章でみたように $1 - \frac{\log 10}{10 \log 4} \sim \frac{9}{10}$ である．

予想 1 画像 $f_N = f : \square_N \to C$ に対して画像 $f_{N-1} : \square_{N-1} \to C$ を

$$f_{N-1}(x) = \left\lfloor \frac{1}{4} \sum_{\square \in \frac{1}{2^{N-1}}\square_1 + x} f(\square) \right\rfloor \quad (x \in \square_{N-1})$$

で定義する．帰納的に $f_{N-i} := (f_{N-i+1})_{N-i}$ と定義する．画像の列 $f_N, f_{N-1}, f_{N-2}, \ldots, f_0$ は次第に画質が荒くなっていくが，画質の良し悪しの変化を認識する境目は $\frac{k}{N} \leq 1 - \frac{\log N}{N \log 4} \leq \frac{k+1}{N}$ を満たすような f_k と f_{k+1} の間である．

5. パーシステントホモロジーによる高次化

パーシステントホモロジーの一般論については [H] が詳しい．depth と 0 次パーステントホモロジーとの関連を見るために，画像 $f : \square_N \to C$ に対して次のようなフィルター付き方体複体 $C_d(f)$ を考える．頂点集合は $\square_N$ であり，2 つの相異なる頂点 a, b はそれらがある $\square \in \square_N^d$ に隣り合って含まれていてかつ $f(a) = f(b)$ であるときに 1 方体で結ばれているとする．2 方体についても同様に定義する．このとき複体 $C_d(f)$ の 0 次ホモロジーの階数は d について広義単調増加関数であるが，あるところから指数的に増加する．その変化の点が depth と対応する．本研究では 0 次パーシステントホモロジーしか考えていないが，同様にして高次の階数から画像の内在的な情報を取り出せると期待できる．

6. 画像認識への応用の展望

機械学習・深層学習において，例えばコンピュータに犬と猫の写真を分類させようとすると，必要な学習データはそれぞれの写真 10000 枚程度とされている．10000 枚のデータを人の手で収集し，それをコンピュータに読み込ませることはかなりのコストを費やすため近年では収集・読み込みの自動化が試みられている．一方で我々の depth を用いた方法は画像の内在的な情報を数理的に引き出すことで画像の分類をしているため学習データを用意する必要がない．未だ精度が荒く実用化への障害はあるものの，パーシステントホモロジーを始めとする様々な数学を用いた画像認識への新たなアプローチとして期待できると考えている．

7. 謝辞

本研究は数物フロンティア・リーディング大学院のプログラムの一つである社会数理実践研究として行われたものである．画像認識についての解説などで尽力してくださった株式会社ニコン研究開発本部数理科学研究所の皆様，特に深層学習との関連や論文に対する貴重なコメントを下さった高山侑也さん，中村ちからさんに心から感謝申し上げます．またセミナーの時間調整や全般に関わるコメントをして頂いた東京大学数理科学研究科特任助教（当時）の土岡俊介さんにも御礼申し上げます．

参考文献

[H] 平岡裕章:「タンパク質構造とトポロジー —パーシステントホモロジー群入門—」共立出版, 2013.