

ニューラル幾何変換場による非剛体画像位置合わせ

石原 実[†] 船富 卓哉[†] 藤村 友貴[†] 久保 尋之^{††} 向川 康博[†]

[†] 奈良先端科学技術大学院大学

^{††} 千葉大学

あらまし 本稿ではニューラル幾何変換場による画像の非剛体位置合わせを提案する。幾何変換場は、幾何変換が空間的に滑らかに変化する場であり、ベクトル場では表現が難しい変形を表現できる。先行研究では、画像間の特徴マッチングをサンプルとして幾何変換場を推定するカーネル回帰手法が提案されているが、回帰を行う際、幾何変換を推定するために対応点のクラスタリングが必要であり、サンプルによっては適切な変換を抽出できない問題がある。その問題に対して提案手法では、カーネル回帰によるモデル化をニューラル場に置き換え、ニューラル幾何変換場を構築した。公開画像データセットを用いて検証を行い、提案手法がカーネル回帰の性能を上回ることを確認した。

キーワード 非剛体位置合わせ, 幾何変換場, ニューラル場, Iterative Blending, Special Euclidean Group

1. はじめに

画像の位置合わせは、2枚の画像間のずれや変形を推定し、同一の位置にある画素同士が重なるように、画像を変形させるタスクである。例えばCTやMRI、生体組織切片画像といった医用画像に対する重要なタスクの一つである。特に、非線形な変形を推定するタスクである非剛体位置合わせについては、精度、計算コスト、汎用性などの向上のために様々な研究がおこなわれてきた。

船富ら[1]は、非剛体変形を表現するモデルとして、幾何変換場を導入した。これは、幾何変換が空間的に滑らかに変化する場であり、従来のベクトル場に比べて複雑な変形場を簡潔に表現できる点で優れている。一方で、[1]では各点における剛体変換をクラスタリングによって計算し、カーネル回帰により回帰を行う手法を用いており、対応点によっては変形場を得られない問題が存在する。

そこで本研究では、対応点を直接学習に用いることができる回帰モデルとしてニューラル場を導入し、幾何変換場のモデル化をクラスタリングなしに実現することを目的とする。

2. カーネル回帰による幾何変換場の推定

船富ら[1]は、局所的な剛体変換（並進+回転）を推定するため、特徴点对を K 個のクラスタに分け、各クラスタから剛体変換 $y_k \in \text{SE}(2)$ (特殊ユークリッド群) を推定した。この変換を各クラスタの特徴点の重心位置 \mathbf{x}_k に対応付け、それら局所的な剛体変換を空間座標から推定するカーネル回帰によって変形場を表現する手法を提案した。

一般的なカーネル回帰は以下で定式化される。

$$y(\mathbf{x}) = \sum_i a_i \phi_i(\mathbf{x}) \quad (1)$$

回帰による変形場のモデリングにおいては、 \mathbf{x} は2次元画像における座標であり、 $\phi_i(\mathbf{x})$ はカーネル関数、 $a_i \in \mathbb{R}$ は各カーネル関数に対応する係数である。一方、 $\text{SE}(2)$ は一般的なカーネル回帰で扱われるスカラー値とは異なり、線形結合で閉じておらず、代数的制約が存在する。そのため、この手法では $\text{SE}(2)$ に対するカーネル回帰を以下で定式化した。

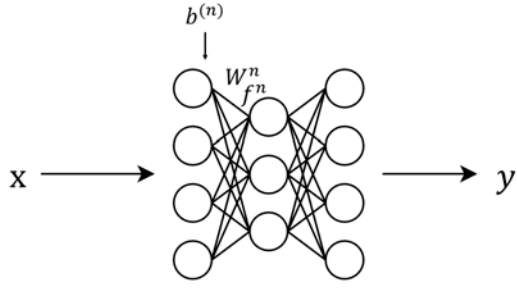
$$y(\mathbf{x}) = \underset{q \in \text{SE}(2)}{\operatorname{argmin}} \left\| \sum_i \frac{\phi_i(\mathbf{x})}{\sum_j \phi_j(\mathbf{x})} \operatorname{Log}_q(c_i) \right\| \quad (2)$$

ここで c_i は、式(1)の係数 a_i に対応する剛体変換であり、以降これを基底と呼ぶ。この定式化に基づき、対数写像を用いた指数重心による合成[2]を適用することで、 $\text{SE}(2)$ の代数的構造を保持したまま、カーネル関数の値に応じて剛体変換の基底を合成している。

一方、[1]では基底 $\mathbf{c} = \{c_i\} \in \text{SE}(2)^K$ を求めるために、訓練サンプルとして画像中の座標 \mathbf{x}_k と剛体変換 y_k の組が必要である。そのため、あらかじめ k -means により対応点をクラスタリングし、クラスタ内の点の対応関係から幾何変換を求めている。しかし、クラスタリングにはクラスタ数など様々なハイパーパラメータが必要であり、かつ対応点がある程度クラスタとしてまとまって抽出されていなければならない。そのため、対応点が少ない場合や疎である場合においてはうまく幾何変換を推定することができず、回帰の精度が低下する問題が存在する。

3. 提案手法

そこで本研究では、カーネル回帰による幾何変換場のモデル化をニューラル場に置き換え、カーネル関数を学習可能なMLPとして学習することにより、クラスタリングに依存しない回帰を実現し、幾何変換場の推定の精度の向上を目指す。船富ら[1]はクラスタリングを用いた前処理によって得た剛体変



$$\begin{aligned} \mathbf{h}^{(1)} &= f^{(1)}(\mathbf{W}^{(1)}\mathbf{x} + \mathbf{b}^{(1)}) \\ \mathbf{h}^{(2)} &= f^{(2)}(\mathbf{W}^{(2)}\mathbf{h}^{(1)} + \mathbf{b}^{(2)}) \\ &\vdots \\ \mathbf{h}^{(n-1)} &= f^{(n-1)}(\mathbf{W}^{(n-1)}\mathbf{h}^{(n-2)} + \mathbf{b}^{(n-1)}) \\ y &= f^{(n)}(\mathbf{W}^{(n)}\mathbf{h}^{(n-1)} + \mathbf{b}^{(n)}) \end{aligned}$$

図 1: 一般的なニューラル場のモデル

換と重心座標の組 (y_k, \mathbf{x}_k) をサンプルとしていたが、本手法では特徴点抽出で得られた 2 画像間の対応点 $(\mathbf{x}_m, \mathbf{x}'_m)$ を直接サンプルとして学習を行い、一方の画像の座標 $\mathbf{x}_m \in \mathbb{R}^2$ に対して幾何変換 $y(\mathbf{x}_m)$ を推論する。幾何変換 $y(\mathbf{x}_m)$ を \mathbf{x}_m に適用し、得られた点 $y(\mathbf{x}_m) \cdot \mathbf{x}_m$ が対応するもう一方の画像の特徴点の座標 \mathbf{x}'_m に一致させるよう、ニューラルネットワークを学習させる。

3.1 MLP を用いたモデルとその問題点

ニューラル場を構築する一般的なモデルとして、図 1 のようなモデルが考えられる。モデルへの入力 $\mathbf{x} \in \mathbb{R}^2$ は座標であり、出力 y は剛体変換 $(SE(2))$ である。しかし、単純な学習を行っても、出力 y が剛体変換となることは保証されない。

そこで本手法では、[1] と同様に、いくつかの幾何変換 $\mathbf{c} \in SE(2)^K$ を基底として持ち、これらの合成重み $\mathbf{w} \in \mathbb{R}^K$ を入力座標に応じて適切に決定する MLP を用いることによって幾何変換場をニューラル表現する。具体的には、MLP で画像の座標 $\mathbf{x} \in \mathbb{R}^2$ から係数 $\mathbf{w} = \{w_i\}$ を推定し、Iterative Blending (IB) アルゴリズム [3] を用いて、 \mathbf{w} と基底 \mathbf{c} から剛体変換 $y(\mathbf{x}) \in SE(2)$ を合成する。

3.2 Iterative Blending による幾何変換の合成

Iterative Blending では係数 \mathbf{w} と \mathbf{c} から剛体変換 q を合成する方法として、 $\sum_i w_i \text{Log}_q(c_i) = 0$ を満たす指数重心 q を反復アルゴリズムで求める。適当な初期値 q_0 を定め (恒等変換など)、各基底 c_i を q_j を接点とした対数写像で接平面空間に射影し (図 2)、その接平面空間上での線形結合 $\alpha \sum_i w_i \text{Log}_{q_j}(c_i)$ を指数関数で $SE(2)$ 上に引き戻した点を次の接点 q_{j+1} とし、 $\sum_i w_i \text{Log}_{q_j}(c_i)$ が閾値 thres 以下になるまで繰り返す手法が用いられる (Algorithm 1)。ただし、 α は減衰率である。この際、 w_i は非負でなければ収束しない可能性がある点に注意が必要である。

3.3 MLP による重みの推定

MLP のアーキテクチャとして、NeRF [4] を参考に、より単純

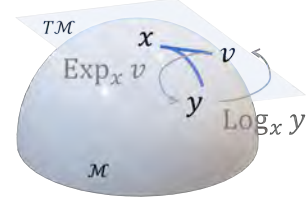


図 2: $SE(2)$ の接平面空間

Algorithm 1 Optimization technique on Lie group $SE(2)$.

Input:

Initial value $q_0 \in SE(2)$, weights $w_i \in \mathbb{R}^+$, basis $c_i \in SE(2)$,
decay rate $\alpha \in \mathbb{R}$, $0 < \alpha \leq 1$

Output: Optimal value $q \in SE(2)$

```
while  $\sum_i w_i \text{Log}_{q_j}(c_i) > \text{thres}$  do
   $q \leftarrow \text{Exp}_q(\alpha \sum_i w_i \text{Log}_{q_j}(c_i))$ 
end while
```

な、全結合層と Leaky ReLU を 4 層重ねたモデルを用いる (図 4)。IB の収束性から $w_i \geq 0$ とするため、出力層の活性化関数は ReLU としている。[4] では入力の座標と出力の画像の周波数の違いから位置埋め込みが用いられていたが、対象となる切片画像における幾何変換場には高周波の成分は含まれていないため、本手法では導入していない。

3.4 ニューラル幾何変換場のモデル

MLP から出力される係数 $\mathbf{w}(\mathbf{x}; \theta) \in \mathbb{R}^K$ と IB アルゴリズムを用いて、最終的な出力 $y(\mathbf{x}) \in SE(2)$ を出力する。

$$\mathbf{w}(\mathbf{x}; \theta) = \{w_i(\mathbf{x}; \theta)\} \in \mathbb{R}^K \quad (3)$$

$$y(\mathbf{x}; \theta, \mathbf{c}) = \text{IB}(\mathbf{w}(\mathbf{x}; \theta), \mathbf{c}) \quad (4)$$

Algorithm 1 では $\sum_i w_i \text{Log}_q(c_i)$ が閾値 thres 以下になるまで q の更新を繰り返すが、最適化における勾配計算で問題が起こる。そのため、実装では Deep Unrolling [5] に基づき、IB の繰り返し回数は固定値として十分な位置合わせ精度が得られる 15 回まで更新を繰り返す。そのモデルが図 3 である。なお簡単のため、モデルパラメータ θ, \mathbf{c} が重要である時を除き、 $y(\mathbf{x}; \theta, \mathbf{c})$ は $y(\mathbf{x})$ と表記する。

3.5 モデルの学習

モデルの学習に用いる損失関数を以下のように定めた。

$$\begin{aligned} \mathcal{L}(\theta, \mathbf{c}) &= \sum_m \|y(\mathbf{x}_m; \theta, \mathbf{c}) \cdot \mathbf{x}_m - \mathbf{x}'_m\|^2 \\ &\quad + \beta \sum_i \sqrt{\sum_m (w_i(\mathbf{x}_m; \theta))^2} \end{aligned} \quad (5)$$

第 1 項は、幾何変換による変換後の座標 $\mathbf{x}' = y(\mathbf{x}_m) \cdot \mathbf{x}_m$ と正解座標 \mathbf{x}'_m の平均二乗誤差である。また、過学習を抑えて汎化性能を担保するため、スパース正則化項を導入し幾何変換

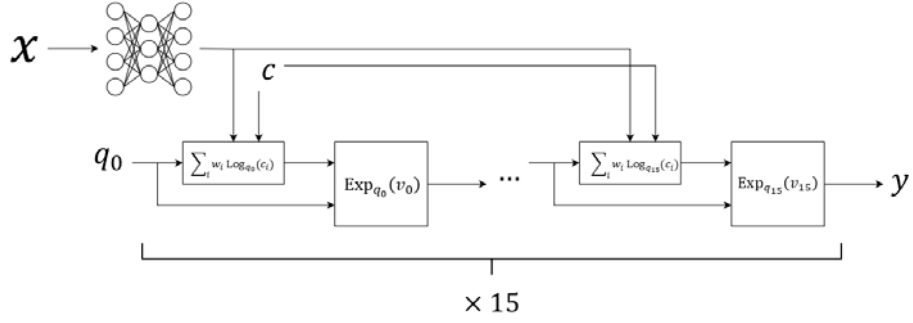


図 3: 提案モデル

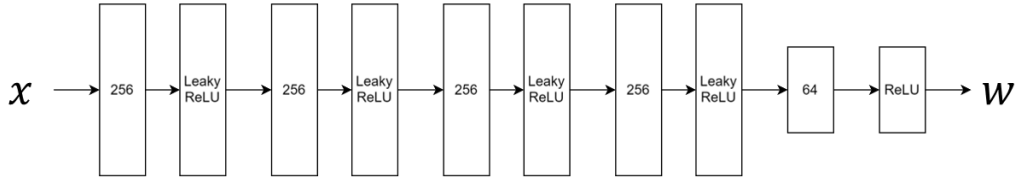


図 4: MLP アーキテクチャ

Algorithm 2 Optimization technique on Lie group SE(2).

Input: Initial value $\{c_i\} \in \text{SE}(2)^K$, Learning rate $\alpha \in \mathbb{R}$,

Loss function $\mathcal{L}(\theta, \mathbf{c}) \in \mathbb{R}$, #iteration $\text{Itr} \in \mathbb{N}$

Output: Optimal value $c_i \in \text{SE}(2)$

for $j = 1$ to Itr **do**

$v \leftarrow 0 \in \mathfrak{se}(2)$

$\Delta v \leftarrow -\alpha \frac{\partial}{\partial v} \mathcal{L}(\theta, \{\text{Exp}_{c_i}(v)\})$

$c_i \leftarrow \text{Exp}_{c_i}(\Delta v)$

end for

場をより少ない基底で表現する。ここで、ある特定の座標 \mathbf{x}_m に対していずれかの $w_i(\mathbf{x}_m; \theta)$ が 0 になるのではなく、いくつかの $w_i(\mathbf{x}_m; \theta)$ が座標全体に対して 0 となるように、訓練データのバッチをパラメータ群として Group Lasso [6] を適用する。

モデルの学習には誤差逆伝播を用いるが、基底である \mathbf{c} を更新する際、SE(2) は加法で閉じていないため、一般的な勾配降下法 (式 (6)) は使えない。

$$c_i \leftarrow c_i - \alpha \frac{\partial}{\partial c_i} \mathcal{L}(\theta, \mathbf{c}) \quad (6)$$

そのため、各基底 c_i を SE(2) の接平面空間に射影し (v)、接平面空間上で更新を行った後 SE(2) の空間に再び射影する方法を用いる (Algorithm 2)。

3.6 実装

モデルの実装には pytorch を使用した。また、Special Euclidean Group SE(2) には、メモリや計算上の優位性から、Anti-Commutative Dual Complex [7] を採用した。重み \mathbf{w} および基底 \mathbf{c} の次元 K は 21 と設定し、スパース正則化によって圧縮表現を獲得することを試みた。

4. 評価実験

4.1 データセット

連続切片画像の公開データセットである ANHIR (Automatic Non-Rigid Histological Image Registration) [8] を用いて評価実験を行った。ANHIR は、顕微鏡で撮影された生物組織の連続切片画像からなる公開データセットであり、組織の変形や歪みを考慮した非剛体レジストレーションの研究に広く利用されている。ANHIR の各画像に対して OmniGlue [9] で特徴点マッチングを行い、1 枚当たり平均約 600 点の対応点が抽出された。本研究では、その対応点を訓練データ $\{\mathbf{x}_m, \mathbf{x}'_m\}$ としてモデルを学習し、評価を行った。

4.2 rTRE による評価

学習によって得られた非剛体位置合わせの変形場に対する評価指標として、 $rTRE$ を用いた。

$$rTRE = \frac{\|y(\mathbf{x}_l) \cdot \mathbf{x}_l - \mathbf{x}'_l\|}{\sqrt{\text{width}^2 + \text{height}^2}} \quad (7)$$

ここで、width は画像の横のピクセル数、height は画像の縦のピクセル数である。 $rTRE$ はモデルの出力座標と正解座標のユークリッド距離を画像の対角線の長さで正規化したもので、画像の大きさに依存しない比較が可能となる。

ANHIR の画像 87 枚に対し、[1] と本研究の比較を $rTRE$ を用いて行った。なお、 $rTRE$ の評価には、ANHIR から提供された対応点 $\{\mathbf{x}_l, \mathbf{x}'_l\}$ を用いた (式 (7))。

図 5 は、各画像に対し、[1] の手法を用いた場合の $rTRE$ を横軸に、本研究の手法を用いた場合の $rTRE$ を縦軸にプロットした図である。 $rTRE$ による定量評価ではほとんどの切片画像で [1] を上回る精度が得られた。一方、図 5 の中央上にあるサ

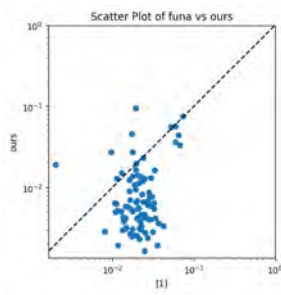
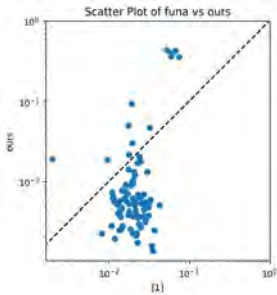


図 5: $rTRE$ による比較

図 6: 手動位置合わせ後の比較

サンプル群は [1] の $rTRE$ の精度を下回った。この切片画像群および画像に対する特徴点抽出の結果を図 7 に示す。これらの画像は、位置合わせ対象の画像全体が 180 度近く回転している画像であった。ニューラル幾何変換場がこれらの大きな変化を表現しきれなかったと仮定し、図 7 の各画像を手動で角度を調整した上で再度 $rTRE$ で評価を行った。その結果が図 6 である。図 5 において中央上にあるサンプル群に対する $rTRE$ は、図 6 では [1] の精度を上回っており、図 7 の画像において位置合わせ精度が低下した原因は、画像全体が大きく回転したことにありと考えられる。

4.3 画像に対する位置合わせ結果

対応点だけでなく、画像全体においてニューラル幾何変換場の精度を確認するため、ニューラル幾何変換場を用いて画像全体を変形させた。その結果が図 8 である。図 8(b) の右図、正解画像と変形後の画像の重ね合わせから分かる通り、変形対象の画像 (Source 画像) に幾何変換場を適用して変形させた結果が、一致させる対象の正解画像 (Target 画像) に限りなく一致しており、画像全体における精度が確かめられた。本手法のニューラル幾何変換場が、幾何変換が空間的に滑らかに変化する場としての性質を満たし、位置合わせにおいて妥当であることがわかる。

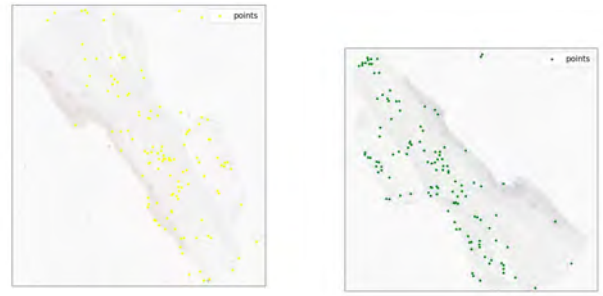
4.4 重み w_i に対するスパース正則化の効果

スパース正則化の効果を検証するため、スパース正則化を導入する場合としない場合で画像の変形結果を比較した (図 9)。また、その際の画像の座標全体に対する重み w_i の値をヒートマップとして可視化した結果を図 10) に示す。スパース正則化を導入した場合には、変形場が w_i の大部分が 0 となっており、スパース正則化の効果が確認できる。

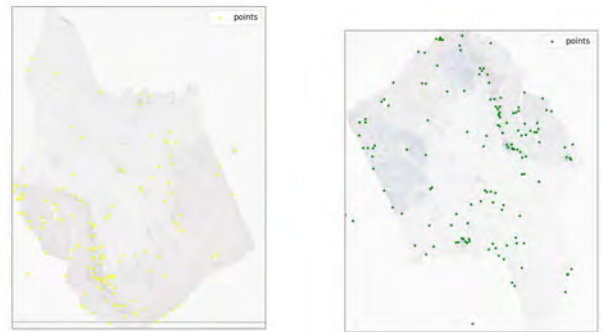
5. 結論および今後の課題

本稿では、幾何変換場による非剛体位置合わせの精度の向上を目的とした、ニューラル幾何変換場の構築について述べた。[1] におけるカーネル回帰による幾何変換場のモデル化をニューラル場に置き換えつつ、Iterative Blending を用いることで剛体変換を合成し、幾何変換場による位置合わせを実現した。また、幾何変換場の基底 c_i に対する重み w_i へのスパース正則化の導入により、少ない基底でのニューラル幾何変換場の表現を可能とした。

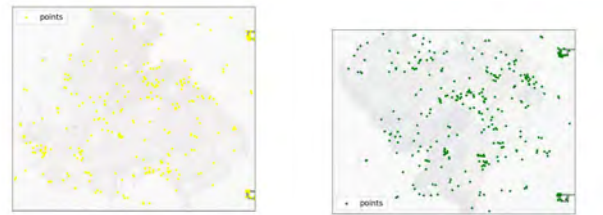
切片画像の公開データセットである ANHIR を用いて [1] と



(a) サンプル (1)



(b) サンプル (2)



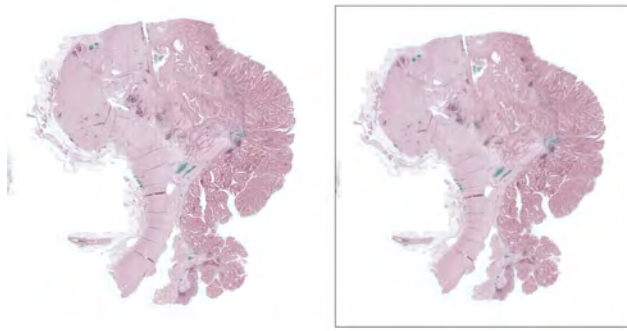
(c) サンプル (3)



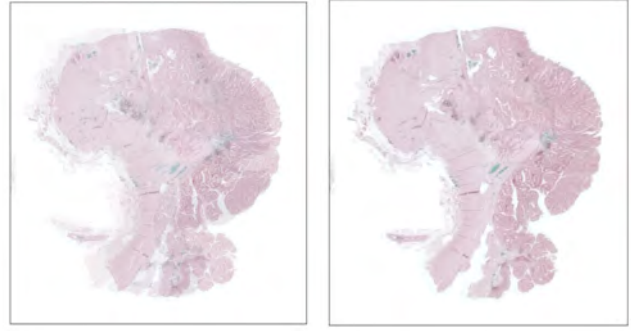
(d) サンプル (4)

図 7: [1] の精度を下回ったサンプル群

の比較を行い、ニューラル幾何変換場が従来モデルの精度を上回ることを確認した。一方、画像全体が大きく移動・回転するケースにおいては [1] より精度で劣っており、前処理として初期位置合わせの導入が考えられる。また、スパース正則化を

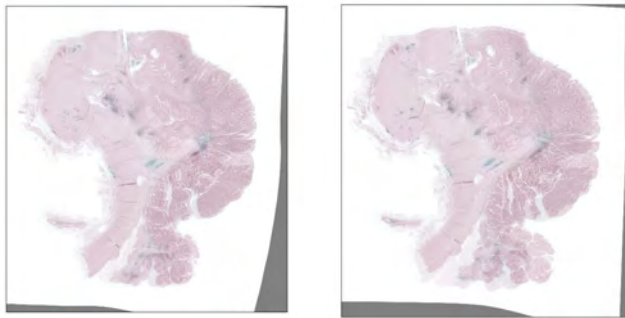


(a) 左: 変形対象の画像 (Source 画像)
右: 変形後の画像



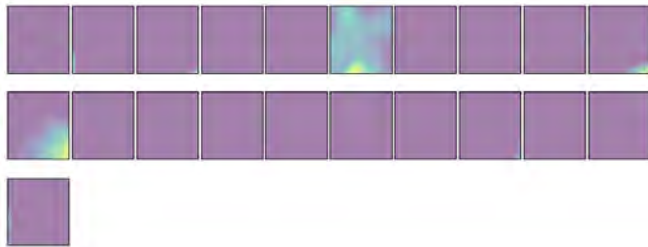
(b) 左: 変形対象の画像 (Source 画像) と変形後の画像
右: 正解画像 (Target 画像) と変形後の画像

図 8: 画像の変形結果

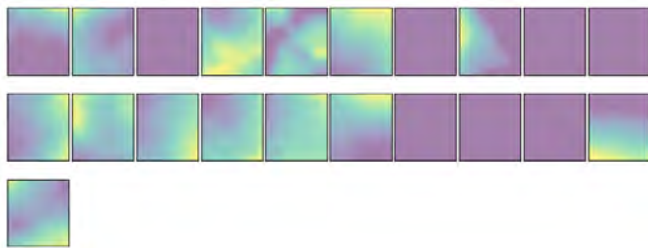


(a) スパース正則化を導入した場合 (b) スパース正則化を導入しない場合

図 9: スパース正則化の有無による画像変形の比較



(a) スパース正則化を導入した場合



(b) スパース正則化を導入しない場合

図 10: 重み w の可視化結果

導入した場合としない場合を比較し、係数 w_i をヒートマップとして可視化することで、スパース正則化の効果を確認した。

ニューラル幾何変換場は非剛体位置合わせに限らず、物体の変形・移動を記述を扱う分野全般への応用が可能である。今後、流体の速度場など、ニューラル幾何変換場を非剛体位置合わせ以外のタスクにも適用し、その有用性を探求していきたい。

謝辞 本研究の一部は JST さきがけ JPMJPR2025, JST ムーンショット型研究開発事業 JPMJMS2389 の支援を受けた。

文 献

- [1] 船富卓哉, 山田重人, 藤村友貴, 櫛田貴弘, 向川康博, “幾何変換のスパース回帰による変形場の表現,” 情処研報 CGVI, pp.184–21, 2021.
- [2] X. Pennec and V. Arsigny, “Exponential barycenters of the canonical cartan connection and invariant means on lie groups,” pp.123–166, Springer Berlin Heidelberg, Berlin, Heidelberg, 2013.
- [3] L. Kavan, S. Collins, J. Zara, and C. O’Sullivan, “Skinning with dual quaternions,” I3D, pp.39–46, 2007.
- [4] B. Mildenhall, P.P. Srinivasan, M. Tancik, J.T. Barron, R. Ramamoorthi, and R. Ng, “NeRF: representing scenes as neural radiance fields for view synthesis,” Commun. ACM, vol.65, no.1, p.99 – 106, Dec. 2021. <https://doi.org/10.1145/3503250>
- [5] V. Monga, Y. Li, and Y.C. Eldar, “Algorithm unrolling: Interpretable, efficient deep learning for signal and image processing,” IEEE Signal Processing Magazine, vol.38, no.2, pp.18–44, 2021.
- [6] M. Yuan and Y. Lin, “Model selection and estimation in regression with grouped variables,” Journal of the Royal Statistical Society: Series B (Statistical Methodology), vol.68, no.1, pp.49–67, 2006. <https://rss.onlinelibrary.wiley.com/doi/abs/10.1111/j.1467-9868.2005.00532.x>
- [7] G. Matsuda, S. Kaji, and H. Ochiai, “Anti-commutative dual complex numbers and 2d rigid transformation,” pp.131–138, Springer Japan, Tokyo, 2014. https://doi.org/10.1007/978-4-431-55007-5_17
- [8] J. Borgev, J. Kybic, I. Arganda-Carreras, D.V. Sorokin, G. Bueno, A.V. Khvostikov, S. Bakas, E.I.-C. Chang, S. Heldmann, K. Kartasalo, L. Latonen, J. Lotz, M. Noga, S. Pati, K. Punithakumar, P. Ruusuvoori, A. Skalski, N. Tahmasebi, M. Valkonen, L. Venet, Y. Wang, N. Weiss, M. Wodzinski, Y. Xiang, Y. Xu, Y. Yan, P. Yushkevich, S. Zhao, and A. Muñoz-Barrutia, “ANHIR: Automatic non-rigid histological image registration challenge,” IEEE Transactions on Medical Imaging, vol.39, no.10, pp.3042–3052, 2020.
- [9] H. Jiang, A. Karpur, B. Cao, Q. Huang, and A. Araujo, “OmniGlue: Generalizable feature matching with foundation model guidance,” 2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), pp.19865–19875, 2024.