

**The supplementary materials of the paper titled “A
Landmark-based Framework for Image Comparison in
the Presence of Various Geometric Misalignments” by
Anik Roy, Sagar Ghosh, and Partha Sarathi Mukherjee**

Contents:

- Section [S.0.1](#): Sketches of the proof of Theorem 3.1.
- Section [S.0.2](#): Theoretical justification of the bootstrap procedure described in Section 2.5.2.
- Section [S.0.3](#): Detection of the locations of maximal change.
- Section [S.0.4](#): Empirical null distribution by the bootstrap procedure described in Section 2.5.2.
- Section [S.0.5](#): Sensitivity analysis of the bootstrap sampling of the proposed method and training sample-size for the autoencoder-based DL method.
- Section [S.0.6](#): Additional numerical study on a synthetic brain image.
- Section [S.0.7](#): Repository link to the **R**-codes. URL: https://github.com/sagarghosh1729/Misaligned_Image_Comparison

S.0.1. Sketches of the proof of Theorem 3.1:

Suppose that the true JLCs of the two given images are $\Gamma^{(1)}$ and $\Gamma^{(2)}$, the sets of detected edge pixels by the LLK-based edge detection method described in Section 2.1 are $\hat{\Gamma}_n^{(1)}$ and $\hat{\Gamma}_n^{(2)}$. Following the line of proof provided by Qiu (1998), we can easily verify that under certain regularity conditions, $\hat{\Gamma}_n^{(\ell)}$ is a strongly consistent estimator of $\Gamma^{(\ell)}$, $\ell = 1, 2$ in Hausdorff distance,

which is defined as:

$$d_H(\widehat{\Gamma}_n^{(\ell)}, \Gamma^{(\ell)}) = \max \left\{ \sup_{(u,v) \in \widehat{\Gamma}_n^{(\ell)}} \inf_{(u',v') \in \Gamma^{(\ell)}} \|(u,v) - (u',v')\|, \sup_{(u,v) \in \Gamma^{(\ell)}} \inf_{(u',v') \in \widehat{\Gamma}_n^{(\ell)}} \|(u,v) - (u',v')\| \right\}. \quad (\text{S.Eq.1})$$

Following similar argument, if $h_n \rightarrow 0$ and other regularity conditions hold,

$$d_H(\widehat{\Gamma}_n^{(\ell)}, \Gamma^{(\ell)}) = \mathcal{O}(h_n) \quad a.s., \quad \ell = 1, 2. \quad (\text{S.Eq.2})$$

Proof of Theorem 3.1: Let X_1 and X_2 be the two landmark matrices for the given two images. Their centering forms are $A = CX_1$ and $B = CX_2$, respectively. Assume that

$$A = CX_1 = \begin{bmatrix} x_1 & y_1 \\ \vdots & \vdots \\ x_{k_n} & y_{k_n} \end{bmatrix} \quad (\text{S.Eq.3})$$

and

$$B = CX_2 = \begin{bmatrix} p_1 & q_1 \\ \vdots & \vdots \\ p_{k_n} & q_{k_n} \end{bmatrix}. \quad (\text{S.Eq.4})$$

Let us first assume that there is no rotation or scaling of the image object, and only translation may have occurred. Under this situation, the result stated in Eqn. (S.Eq.2) ensures

$$\|(x_i, y_i) - (p_i, q_i)\| \leq \sqrt{\theta_n^2 + h_n^2} = \delta \quad a.s., \quad \text{for all } i \in \{1, 2, \dots, k_n\}. \quad (\text{S.Eq.5})$$

Since we know that on a finite-dimensional normed linear space, any two

norms are equivalent [Theorem, 2.4-5, Kreyszig (1991)], $\|\cdot\|_{L_\infty}$ and $\|\cdot\|_F$ are also equivalent. Since Eqn. (S.Eq.5) is true for every pair of (x_i, y_i) and (p_i, q_i) , we have $\|A - B\|_{L_\infty} \leq \delta$ a.s.

Now, following the notation that $X \leq Y \implies (Y - X)$ is a matrix consisting of non-negative entries, we get

$$B - \delta J \leq A \leq B + \delta J, \quad (\text{S.Eq.6})$$

where J is a $k_n \times 2$ matrix whose all entries are 1. This leads us to the following inequality

$$AA^T \leq BB^T + \mathcal{O}(\delta)JB^T + \mathcal{O}(\delta)BJ^T + \mathcal{O}(\delta^2)JJ^T. \quad (\text{S.Eq.7})$$

From Eqn. (S.Eq.5), we have

$$|x_i - p_i| \leq \delta \text{ and } |y_i - q_i| \leq \delta, \text{ for all } i \in \{1, 2, \dots, k_n\}. \quad (\text{S.Eq.8})$$

By using the triangle inequality repeatedly, we get

$$\begin{aligned} \left| (AA^T)_{ij} - (BB^T)_{ij} \right| &= |(x_i x_j + y_i y_j) - (p_i p_j + q_i q_j)| \\ &\leq |x_i x_j - p_i p_j| + |y_i y_j - q_i q_j| \\ &= |x_i x_j - x_i p_j + x_i p_j - p_i p_j| + |y_i y_j - y_i q_j + y_i q_j - q_i q_j| \\ &\leq |x_i| |x_j - p_j| + |p_j| |x_i - p_i| + |y_i| |y_j - q_j| + |q_j| |y_i - q_i| \\ &\leq 4\delta = \mathcal{O}(\delta) = \mathcal{O}(h_n). \end{aligned} \quad (\text{S.Eq.9})$$

Therefore, we have

$$\left| \|AA^T\|_{L_\infty} - \|BB^T\|_{L_\infty} \right| \leq \|AA^T - BB^T\|_{L_\infty} = \mathcal{O}(\delta). \quad (\text{S.Eq.10})$$

Since $h_n \rightarrow 0$ as $n \rightarrow \infty$, there exists $n_0 \in \mathbb{N}$ such that for all $n \geq n_0$,

$$\frac{1}{1 - \mathcal{O}(h_n)} \leq 1 + 2\mathcal{O}(h_n). \quad (\text{S.Eq.11})$$

Let $a = \|AA^T\|_{L_\infty}$ and $b = \|BB^T\|_{L_\infty}$. Using the previous bounds and Eqn. (S.Eq.7), we get

$$\begin{aligned} \frac{AA^T}{a} &\leq \frac{BB^T + \mathcal{O}(\delta)JB^T + \mathcal{O}(\delta)BJ^T + \mathcal{O}(\delta^2)JJ^T}{(b - \mathcal{O}(\delta))} \\ &= \frac{BB^T + \mathcal{O}(h_n)JB^T + \mathcal{O}(h_n)BJ^T + \mathcal{O}(h_n^2)JJ^T}{(b - \mathcal{O}(h_n))} \\ &= \frac{BB^T + \mathcal{O}(h_n)JB^T + \mathcal{O}(h_n)BJ^T + \mathcal{O}(h_n^2)JJ^T}{b(1 - \mathcal{O}(h_n))} \\ &\leq \frac{BB^T + \mathcal{O}(h_n)JB^T + \mathcal{O}(h_n)BJ^T + \mathcal{O}(h_n^2)JJ^T}{b} (1 + 2\mathcal{O}(h_n)) \\ &\leq \frac{BB^T}{b} + \mathcal{O}(h_n)Z, \end{aligned} \quad (\text{S.Eq.12})$$

where

$$Z := \frac{1}{b} \left[JB^T + BJ^T + \mathcal{O}(h_n) \{ JJ^T + 2BB^T + 2JB^T + 2BJ^T + \mathcal{O}(h_n)JJ^T \} \right].$$

Since it is easy to verify that $\|Z\|_{L_\infty} = \mathcal{O}(1)$, we get

$$\left\| \frac{AA^T}{a} - \frac{BB^T}{b} \right\|_{L_\infty} \leq \mathcal{O}(h_n) \|Z\|_{L_\infty} = \mathcal{O}(h_n) \text{ a.s.}, \quad (\text{S.Eq.13})$$

From the previous arguments, we have

$$\left\| \frac{AA^T}{\|AA^T\|_F} - \frac{BB^T}{\|BB^T\|_F} \right\|_{L_\infty} = \mathcal{O}(h_n) \text{ a.s.} \quad (\text{S.Eq.14})$$

Now we incorporate rotation and scaling of the image object in the problem. Consider $E = \alpha BR$, where $\alpha > 0$ is the scaling factor, and $R \in SO(2)$

is a rotation matrix. We note that $EE^T = \alpha^2 BB^T$ as $RR^T = I_2$. Hence,

$$\left\| \frac{AA^T}{\|AA^T\|_F} - \frac{EE^T}{\|EE^T\|_F} \right\|_{L_\infty} = \left\| \frac{AA^T}{\|AA^T\|_F} - \frac{BB^T}{\|BB^T\|_F} \right\|_{L_\infty} = \mathcal{O}(h_n) \text{ a.s.} \quad (\text{S.Eq.15})$$

This completes the proof of $T_{H_0}(W^{(1)}, W^{(2)}) \rightarrow 0$ almost surely, as $n \rightarrow \infty$.

Next, we prove the second statement of Theorem 1 regarding the alternate hypothesis. The proof is similar once we observe that there exists $i \in \{1, 2, \dots, k_n\}$ such that

$$\inf_{\gamma > 0, U \in SO(2)} \left| \|\gamma(AU)_{i*}\| - \|B_{i*}\| \right| > \epsilon \text{ a.s.} \quad (\text{S.Eq.16})$$

for some $\epsilon > 0$, where $(AU)_{i*}$ and B_{i*} denote i -th rows of AU and B , respectively. The above statement holds because there has to be at least one landmark point in the first image, which is at least ϵ distance apart from the nearest landmark point in the second image, up to a rotation transformation. Consider $U = I_2$, and select $i \in \{1, 2, \dots, k_n\}$ such that

$$\inf_{\gamma > 0} \left| \|\gamma A_{i*}\| - \|B_{i*}\| \right| > \epsilon \text{ a.s.} \quad (\text{S.Eq.17})$$

Therefore, we can state that for any $\gamma > 0$,

$$\left| \gamma \sqrt{x_i^2 + y_i^2} - \sqrt{p_i^2 + q_i^2} \right| > \epsilon \text{ a.s.} \implies \left| \gamma^2(x_i^2 + y_i^2) - (p_i^2 + q_i^2) \right| > \epsilon^2 \text{ a.s.} \quad (\text{S.Eq.18})$$

From Eqn. (S.Eq.16), it is easy to check that

$$\begin{aligned}\|\gamma^2 AA^T - BB^T\|_{L_\infty} &= \sup_{\ell, j \in \{1, 2, \dots, k_n\}} \left| \gamma^2 (x_\ell x_j + y_\ell y_j) - (p_\ell p_j + q_\ell q_j) \right| \\ &\geq \left| \gamma^2 (x_i^2 + y_i^2) - (p_i^2 + q_i^2) \right| > \epsilon^2 \text{ a.s.}\end{aligned}\quad (\text{S.Eq.19})$$

The last equation holds for any $\gamma > 0$, in particular for $\gamma^2 := \frac{\|BB^T\|_F}{\|AA^T\|_F}$. Therefore, we get

$$\left\| \frac{AA^T}{\|AA^T\|_F} - \frac{BB^T}{\|BB^T\|_F} \right\|_{L_\infty} = \frac{1}{\|BB^T\|_F} \|\gamma^2 AA^T - BB^T\| > \frac{\epsilon^2}{\|BB^T\|_F} \text{ a.s.} \quad (\text{S.Eq.20})$$

This concludes the proof of the second statement of Theorem 3.1.

S.0.2. Theoretical justification of the bootstrap procedure described in Section 2.5.2

In this section, we study the statistical consistency of the bootstrap procedure described in Section 2.5.2. We first introduce the notation used throughout. Let $\hat{F}_B^*$ denote the empirical distribution function of the bootstrap test statistic generated under H_0 , based on B bootstrap replicates, and let F denote the corresponding true distribution function under H_0 . Our goal is to establish the almost sure uniform convergence

$$\sup_x \left| \hat{F}_B^*(x) - F(x) \right| \xrightarrow{\text{a.s.}} 0, \quad \text{as } B \rightarrow \infty.$$

Notations: Let $\mathcal{I}$ denote the space of binary image objects (obtained after gradient-based thresholding), and let G be the group of transformations consisting of rotations, translations, and scalings acting on $\mathcal{I}$. For any $I \in \mathcal{I}$,

define its equivalence class under G by

$$[I] := \{g \cdot I : g \in G\}.$$

Let $\mathcal{X}$ denote the space of landmark representations associated with elements of $\mathcal{I}$. We write $D : \mathcal{I} \times \mathcal{I} \rightarrow [0, \infty)$ for a shape distance on $\mathcal{I}$ that is invariant to rotation, translation, and scaling, and let $d_M : \mathcal{X} \times \mathcal{X} \rightarrow [0, \infty)$ denote the proposed landmark-based metric. Let $I_1, I_2 \in \mathcal{I}$ be two thresholded binary images, and let $X_1, X_2 \in \mathcal{X}$ be their corresponding landmark matrices. The proposed test statistic is

$$T := D(I_1, I_2) = d_M(X_1, X_2).$$

Under the null hypothesis H_0 (*i.e.*, no shape difference), there exists an unknown template shape $S \in \mathcal{I}$ such that

$$I_1 = g_1 \cdot S, \quad I_2 = g_2 \cdot S,$$

for some unobserved transformations $g_1, g_2 \in G$. We assume g_1 and g_2 are i.i.d. draws from a probability measure μ on G .

Assumptions: (I) $g_1, g_2 \stackrel{\text{i.i.d.}}{\sim} \mu$, where μ is a probability measure on the group G , (II) (*Invariance of μ .*) For all measurable sets $A \subseteq G$ and all $h \in G$,

$$\mu(hA) = \mu(A) \quad \text{and} \quad \mu(h^{-1}Ah) = \mu(A).$$

Bootstrap procedure described in Section 2.5.2: Without loss of generality, consider I_1 as the reference image. Generate B i.i.d. random transformations $g_1^*, \dots, g_B^* \stackrel{\text{i.i.d.}}{\sim} \mu$ on G . For each $b = 1, \dots, B$, construct the b -th bootstrap image by

$$I^{*(b)} := g_b^* \cdot I_1.$$

Let X_1 denote the landmark matrix associated with I_1 , and let $X^{*(b)}$ be the landmark matrix associated with $I^{*(b)}$. Compute the bootstrap replicate of the test statistic as

$$T^{*(b)} := D(I_1, I^{*(b)}) = d_M(X_1, X^{*(b)}).$$

Finally, the empirical distribution of $\{T^{*(b)}\}_{b=1}^B$ is used to approximate the null distribution of the test statistic under H_0 .

Consistency: Under H_0 , we know that, $I_1 = g_1 \cdot S$ and $I_2 = g_2 \cdot S$ for some $g_1, g_2 \in G$. Using the invariance of D under the group action, for any fixed $h \in G$,

$$D(I_1, I_2) = D(g_1 \cdot S, g_2 \cdot S) = D(hg_1 \cdot S, hg_2 \cdot S).$$

Choosing $h = g_1^{-1}$ yields

$$T := D(I_1, I_2) = D(S, g_1^{-1}g_2 \cdot S) = D(S, g \cdot S), \quad g := g_1^{-1}g_2.$$

By the assumed invariance properties of μ , we have $g \sim \mu$.

For the bootstrap, recall that $I^{*(b)} = g_b^* \cdot I_1$ with $g_b^* \stackrel{\text{i.i.d.}}{\sim} \mu$. Then

$$T^{*(b)} := D(I_1, I^{*(b)}) = D(g_1 \cdot S, g_b^* g_1 \cdot S) = D(S, g_1^{-1} g_b^* g_1 \cdot S) = D(S, g^{*(b)} \cdot S),$$

where $g^{*(b)} := g_1^{-1} g_b^* g_1$. Again, by invariance of μ , $g^{*(b)} \sim \mu$, and hence $T^{*(b)} \stackrel{d}{=} T$ under H_0 . Moreover, $\{T^{*(b)}\}_{b=1}^B$ are i.i.d. with common distribution function F . Therefore, letting $\hat{F}_B^*$ denote the empirical distribution function of $\{T^{*(b)}\}_{b=1}^B$, the Glivenko-Cantelli theorem implies

$$\sup_x \left| \hat{F}_B^*(x) - F(x) \right| \xrightarrow{\text{a.s.}} 0, \quad \text{as } B \rightarrow \infty.$$

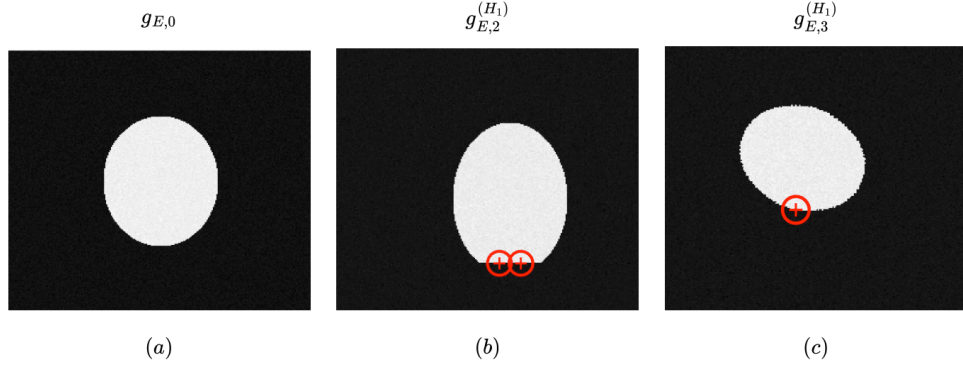


Figure S.1: Locations of maximal changes: (b) shows the regions (two regions with the same maximal changes) in $g_{E,2}^{(H_1)}$ with the largest change from (a), and (c) shows the region in $g_{E,3}^{(H_1)}$ with the largest change from (a).

S.0.3. Detection of the locations of maximal change

Building on Remark 2.5, we perform experiments to identify the locations exhibiting the largest changes in the JLCs. Figure S.1 presents the original ellipse alongside its two changed versions, emphasizing the regions where the most pronounced changes occur. In particular, the second panel of Figure S.1 highlights, with two red circles, the regions that exhibit the greatest deviation from the original object in the first panel of Figure S.1, while the

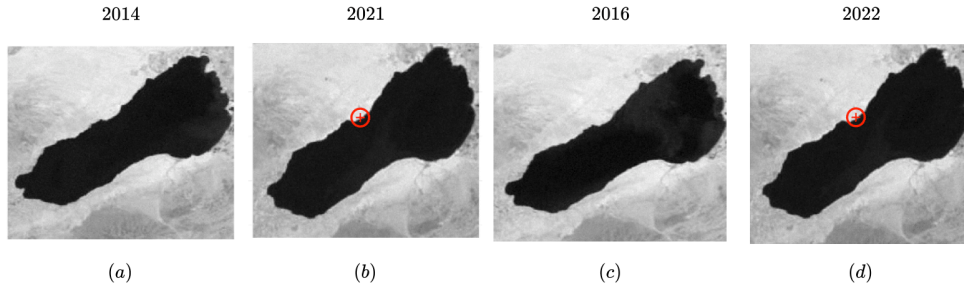


Figure S.2: Locations of maximal changes: (b) shows the region in 2021 with the largest change since 2014 (a), and (d) shows the region in 2022 with the largest change since 2016 (c).

third panel of Figure S.1 shows, again marked with a red circle, the region corresponding to the largest change relative to the original ellipse in the first panel of Figure S.1.

In Figure S.2, we compare the images from 2014 vs. 2021, and 2016 vs. 2022, highlighting the regions of maximal change in each pair. Specifically, in the second panel of Figure S.2, the red circle marks the location in the 2021 image where the largest change occurred since 2014, whereas in the fourth panel of Figure S.2, the red circle indicates the location in the 2022 image corresponding to the maximal change since 2016.

S.0.4. Empirical null distribution by the bootstrap procedure described in Section 2.5.2

Figure S.3 displays the empirical null distributions obtained from 2000 bootstrap samples for the simulated ellipse, convex polygon, and amoeba-shaped image objects.

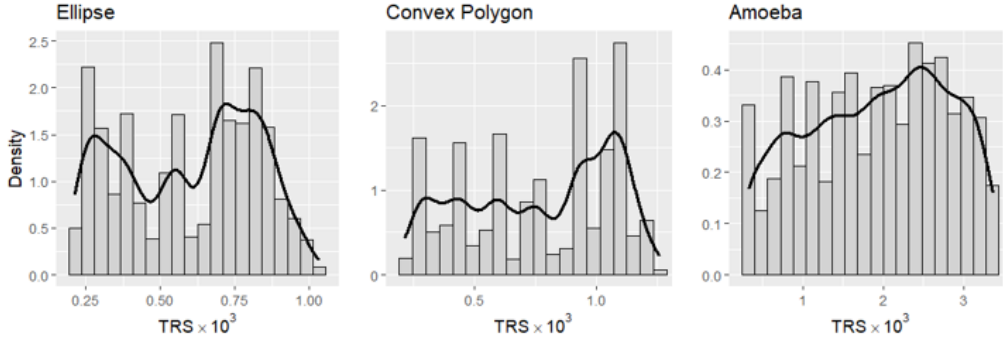


Figure S.3: Visualizations of the estimated densities of the empirical null distributions generated by the bootstrap procedure for the ellipse, convex polygon, and amoeba settings. The X-axis represents $TRS \times 10^3$, while the Y-axis shows the corresponding density. The solid curves represent smoothed density estimates.

S.0.5. Sensitivity analysis of the bootstrap sampling of the proposed method and training sample-size for the autoencoder-based DL method

To assess the stability of the bootstrap procedure described in Section 2.5.2, we evaluate the performance of the proposed method under increasing numbers of bootstrap samples using the amoeba-shaped simulated images. Table S.1 presents the estimated cutoff values for different bootstrap sample

Proposed Method				
Amoeba Shape ($g_{A,0}$)	$B = 200$	$B = 500$	$B = 1000$	$B = 2000$
cutoff	0.003198	0.003190	0.003203	0.003203

Table S.1: Estimated cutoff values under increasing bootstrap sample sizes.

sizes. The results show that the cutoff values remain largely stable as the number of bootstrap samples increases, suggesting that 500 bootstrap samples are adequate in this setting. Table S.2 presents the empirical size and

Comparison	Proposed Method				DL: Autoencoder			
	$B = 200$	500	1000	2000	$Tr = 200$	500	1000	2000
$g_{A,0}$ vs. $g_{A,1}^{(H_0)}$	0.038	0.054	0.032	0.032	0.055	0.057	0.026	0.010
$g_{A,0}$ vs. $g_{A,2}^{(H_1)}$	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
$g_{A,0}$ vs. $g_{A,3}^{(H_1)}$	0.640	0.642	0.638	0.638	0.360	0.764	0.880	0.916

Table S.2: Empirical size (the first row) and power (the second and third rows) of the proposed method for amoeba-shaped image objects across different bootstrap sample sizes. The right side of the table summarizes the performance of the DL method under increasing training sample sizes. For the proposed method, B represents the bootstrap sample size, and for the DL method, Tr represents the training sample size.

power of the proposed method across several amoeba-shaped image settings. The findings suggest that increasing the number of bootstrap samples provides little additional gain in power. In contrast, the deep-learning-based method exhibits improved power as the training sample size grows, which is intuitively expected. At the same time, its empirical size decreases as the amount of training data increases, which may indicate potential overfitting. The observations with the autoencoder-based method are consistent with the

modern over-parametrized models. As the number of training samples grows, the neural network-based architecture may tend to fit the null distribution increasingly well and begin to memorize noise, leading to overfitting.

S.0.6. Additional numerical study on a synthetic brain image

In this example, we show the performance of the proposed testing method on synthetic brain images of a patient captured at various time points. As pre-processing steps, we convert the original color images to grayscale images, and resize each image to the resolution 128×128 . To further emphasize the tumor region, we subsequently apply image segmentation. The panels in Figure S.4 show the tumor growth of the patient (Stamatakis and Giatili, 2017). The images exhibit only minor translation and rotation of the object

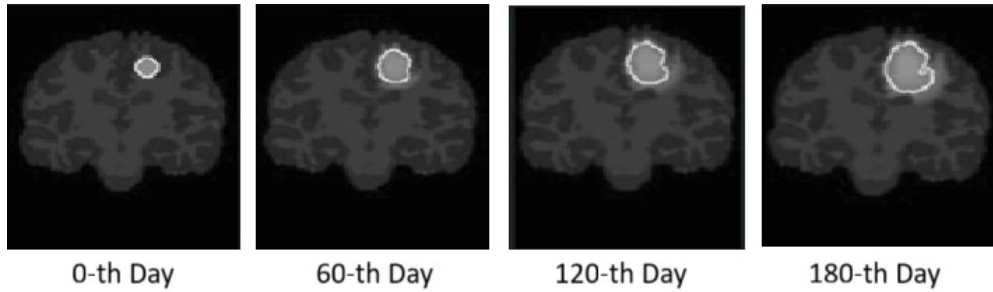


Figure S.4: A virtual glioblastoma tumor growth progression in vivo on a coronal slice at various time points.

and, notably, no change in scale. In this example, our primary interest lies in comparing both the size and the shape of the tumor across time points; thus, we employ the test statistic defined in Eqn. 11. In this study, we first apply the proposed testing method to compare the images captured on the 0-th day and 60-th day. Then, we compare the images captured on the 60-th day and 120-th day, and finally, between the images captured on the 120-th day and 180-th day. Additionally, we apply all competing methods to these data and summarize their performance in Table S.3. From this table, we see that the proposed method is capable of detecting the change in shape of the

tumor nearly perfectly in all the scenarios mentioned above. Furthermore, to assess the control of false rejections, we first apply the proposed test to a purely rotated version of the 0-th day image and obtain a very large p-value, indicating no spurious change detection. This statistical inference is consistent with the visual impression and ground truth, suggesting that both the shape and the size of the tumor have indeed changed in the cases discussed above.

Comparison	Proposed	SD-1	SD-2	SD-3	DL	NIC
0-th vs. rotated 0-th day	0.762	0.100	0.560	0.096	0.364	0.692
0-th vs. 60-th day	0.000	0.000	0.000	0.074	0.010	0.006
60-th vs. 120-th day	0.000	0.092	0.104	0.086	0.008	0.092
120-th vs. 180-th day	0.000	0.020	0.000	0.130	0.066	0.154

Table S.3: p-values of the proposed test for various real image comparisons. SD-1: *convexity*, SD-2: *eccentricity*, SD-3: *total absolute curvature*, DL: autoencoder-based image comparison using deep learning, NIC: nonparametric image comparison.

*S.0.7. Repository link to the **R**-codes*

The **R**-codes are available in the following GitHub repository:

https://github.com/sagarghosh1729/Misaligned_Image_Comparison

References:

- Kreyszig, E. (1991). Introductory functional analysis with applications, volume 17. John Wiley & Sons.
- Qiu, P. (1998). Discontinuous regression surfaces fitting. The Annals of Statistics, 26(6): 2218-2245.
- Stamatakis, G. and Giatili, S. (2017). A numerical handling of the boundary conditions imposed by the skull on an inhomogeneous diffusion-reaction model of glioblastoma invasion into the brain: Clinical validation aspects. Cancer informatics, 16:1-16.