CABLD: Contrast-Agnostic Brain Landmark Detection with Consistency-Based Regularization

Supplementary Material

1. Analytical and Differentiable Coordinate Transformations

Notation: Lowercase bold letters denote column vectors, while uppercase bold letters are used for matrices. Coordinates in D dimensions are represented as column vectors, i.e., $\mathbf{x} \in \mathbb{R}^D$. The symbol $\tilde{\mathbf{x}}$ denotes $\mathbf{x}$ in homogeneous coordinates, expressed as $\tilde{\mathbf{x}} = [\mathbf{x}, 1]^T$. Superscripts like $\mathbf{x}^{(j)}$ are used to indicate distinct instances of $\mathbf{x}$ (such as in a dataset), while subscripts, $\mathbf{x}_j$, represent the j-th component of $\mathbf{x}$.

We introduce families of parametric transformations that can be derived explicitly in closed-form based on corresponding landmark pairs. Let us consider N matching landmark pairs $\{(\mathbf{x}^{(j)}, \mathbf{y}^{(j)})\}_{j=1}^N$, where $\mathbf{x}^{(j)}, \mathbf{y}^{(j)} \in \mathbb{R}^D$ and $N > D$. For simplicity, we define $\mathbf{X} := \langle \mathbf{x}^{(1)} \dots \mathbf{x}^{(N)} \rangle \in \mathbb{R}^{D \times N}$, and similarly for $\tilde{\mathbf{X}}$ and $\mathbf{Y}$. We define a transformation function $T_\beta : \mathbb{R}^D \to \mathbb{R}^D$, where $\beta \in \mathcal{B}$ are the transformation parameters.

1.1. Thin-Plate Spline Deformation Model

The thin-plate spline (TPS) transformation is used for coordinate mapping, delivering a non-rigid, parameterized deformation model with a closed-form solution for interpolating corresponding landmarks [8, 14, 45, 73]. This approach offers greater adaptability than affine mappings while inherently encompassing affine transformations as a specific case.

The TPS deformation model $T_\beta : \mathbb{R}^D \to \mathbb{R}^D$ is expressed as:

$$T_\beta(\mathbf{x}) = \mathbf{W}^T \tilde{\mathbf{x}} + \sum_{j=1}^{N} \mathbf{v}_j \Phi(\|\mathbf{x}^{(j)} - \mathbf{x}\|^2), \tag{S1}$$

where $\mathbf{W} \in \mathbb{R}^{D \times (D+1)}$ and $\mathbf{v}_j \in \mathbb{R}^D$ represent the transformation parameters (β), and $\Phi(r) = r^2 \ln(r)$. Additionally, $\mathbf{V} = \{\mathbf{v}_j\}_{j=1}^N$, making the full parameter set $\beta = \{\mathbf{W}, \mathbf{V}\}$.
The transformation T minimizes the bending energy:

$$I_T = \int_{\mathbb{R}^D} \|\nabla^2 T(\mathbf{x})\|_F^2 \, d\mathbf{x}, \tag{S2}$$

which ensures that T is smooth with square-integrable second derivatives. We impose the interpolation conditions $T(\mathbf{x}^{(j)}) = \mathbf{y}^{(j)}$ for $j = 1, \dots, N$, and the following constraints to ensure a well-posed solution:

$$\sum_{j=1}^{N} \mathbf{v}_j = \mathbf{0} \quad \text{and} \quad \sum_{j=1}^{N} \mathbf{v}_j (\mathbf{x}^{(j)})^T = \mathbf{0}. \tag{S3}$$

Based on the mentioned conditions, the linear system below can be considered for β:

$$\begin{bmatrix} \mathbf{M} & \mathbf{R} \\ \mathbf{R}^T & \mathbf{Z} \end{bmatrix} \begin{bmatrix} \mathbf{V} \\ \mathbf{W} \end{bmatrix} = \begin{bmatrix} \mathbf{Y} \\ \mathbf{Z} \end{bmatrix}, \tag{S4}$$

where $\mathbf{M} \in \mathbb{R}^{N \times N}$ with entries $M_{ij} = \Phi(\|\mathbf{x}^{(i)} - \mathbf{x}^{(j)}\|^2)$, $\mathbf{R} \in \mathbb{R}^{N \times (D+1)}$ where each Row j is $\tilde{\mathbf{x}}^{(j)T}$, $\mathbf{V} \in \mathbb{R}^{N \times D}$ with the j^{th} row being $\mathbf{v}_j^T$, $\mathbf{Y} \in \mathbb{R}^{N \times D}$ with row entries of $\mathbf{y}^{(j)T}$, and $\mathbf{Z}$ is a zero matrix with the proper size.
Accordingly, the solution β^* is obtained by:

$$\beta^* = \begin{bmatrix} \mathbf{V}^* \\ \mathbf{W}^* \end{bmatrix} = \begin{bmatrix} \mathbf{M} & \mathbf{R} \\ \mathbf{R}^T & \mathbf{Z} \end{bmatrix}^{-1} \begin{bmatrix} \mathbf{Y} \\ \mathbf{Z} \end{bmatrix}. \tag{S5}$$

Using the equations above, β^* can be formulated as a differentiable function, ensuring integration with gradient-based optimization frameworks.
Finally, the general TPS equation can be improved (e.g., to handle noise) by incorporating a regularization term:

$$\beta^* = \arg\min_\beta \sum_{j=1}^{N} \|T_\beta(\mathbf{x}^{(j)}) - \mathbf{y}^{(j)}\|^2 + \lambda I_T, \tag{S6}$$

where λ is a positive hyperparameter that determines the regularization level. As $\lambda \to \infty$, the optimal transformation T tends to an affine form. This can be achieved by modifying the matrix $\mathbf{M}$ to $\mathbf{M} + \lambda \mathbf{I}$ in the linear system (Eq. S4). The parameter λ could influence the solution β^*, leading it either toward an affine transformation as $\lambda \to \infty$ or toward a fully nonlinear deformation as $\lambda \to 0$.

2. Random Convolution-Based Contrast Augmentation

Figure S1 illustrates the model architecture used for random convolution (RC)-based contrast augmentation. The model consists of five non-linear blocks, each comprising an RC layer followed by a LeakyReLU activation. This cascaded design efficiently captures complex and non-linear intensity relationships across various MRI contrasts, generating diverse artificial contrast variations from a single scan. Additionally, Figure S2 presents axial mid-slices of augmented

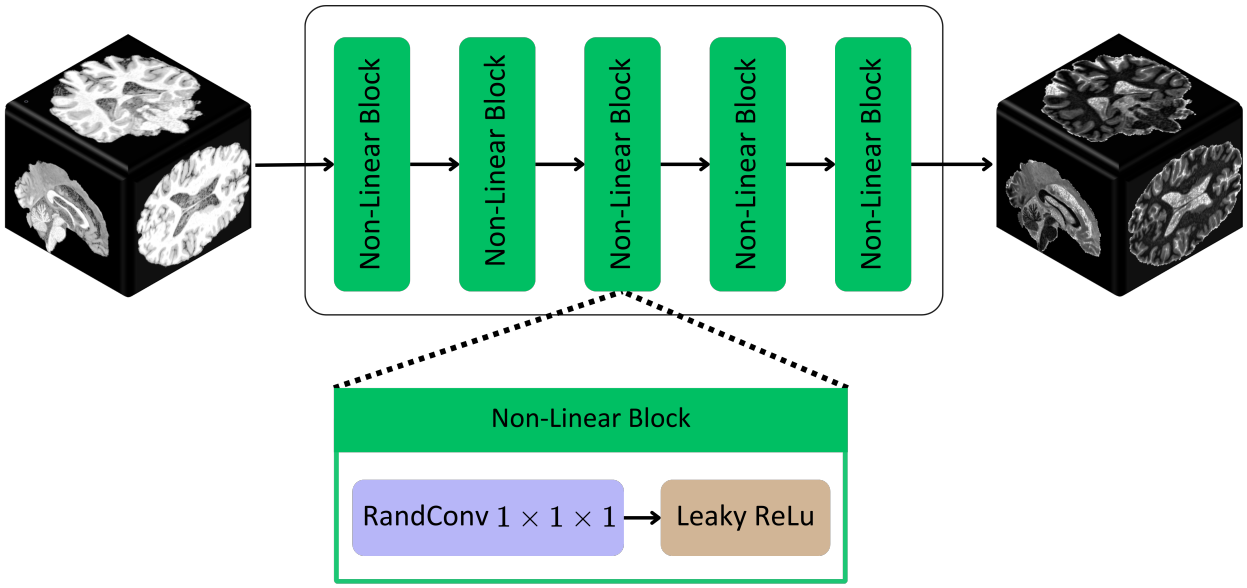

Figure S1. The overall architecture of the proposed 3D contrast augmentation method using random convolution.

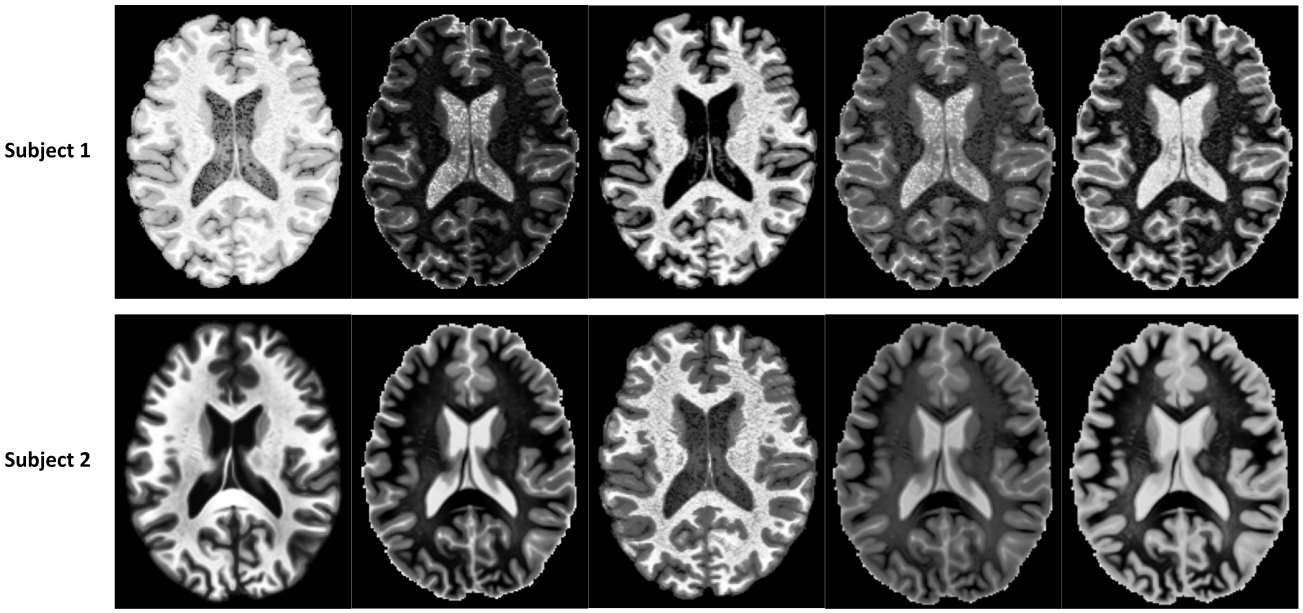

Figure S2. Axial mid-slices of augmented samples generated using the RC-based contrast augmentation method with $1 \times 1 \times 1$ convolution kernels.

samples generated using the RC-based contrast augmentation scheme. These samples demonstrate the effectiveness of RC in simulating a wide range of artificial contrasts from a single input scan.

To investigate the kernel size's impact on the RC output for contrast augmentation, we have implemented the kernel size of $3 \times 3 \times 3$ and $5 \times 5 \times 5$ in all non-linear blocks of our model. Figure S3 showcases axial mid-slices of augmented samples produced using the RC-based contrast augmentation method, incorporating $3 \times 3 \times 3$ and $5 \times 5 \times 5$ convolutions. Evidently, in these samples, the augmented outputs exhibit a noticeable blurring effect, which can adversely affect the performance of the DL models. In particular, this blurring compromises precise voxel-to-voxel correspondences for our task, thereby degrading the accuracy of anatomical landmark detection outcomes.

It is important to note that while the augmented scans with RC are inputs to the anatomical landmark detection

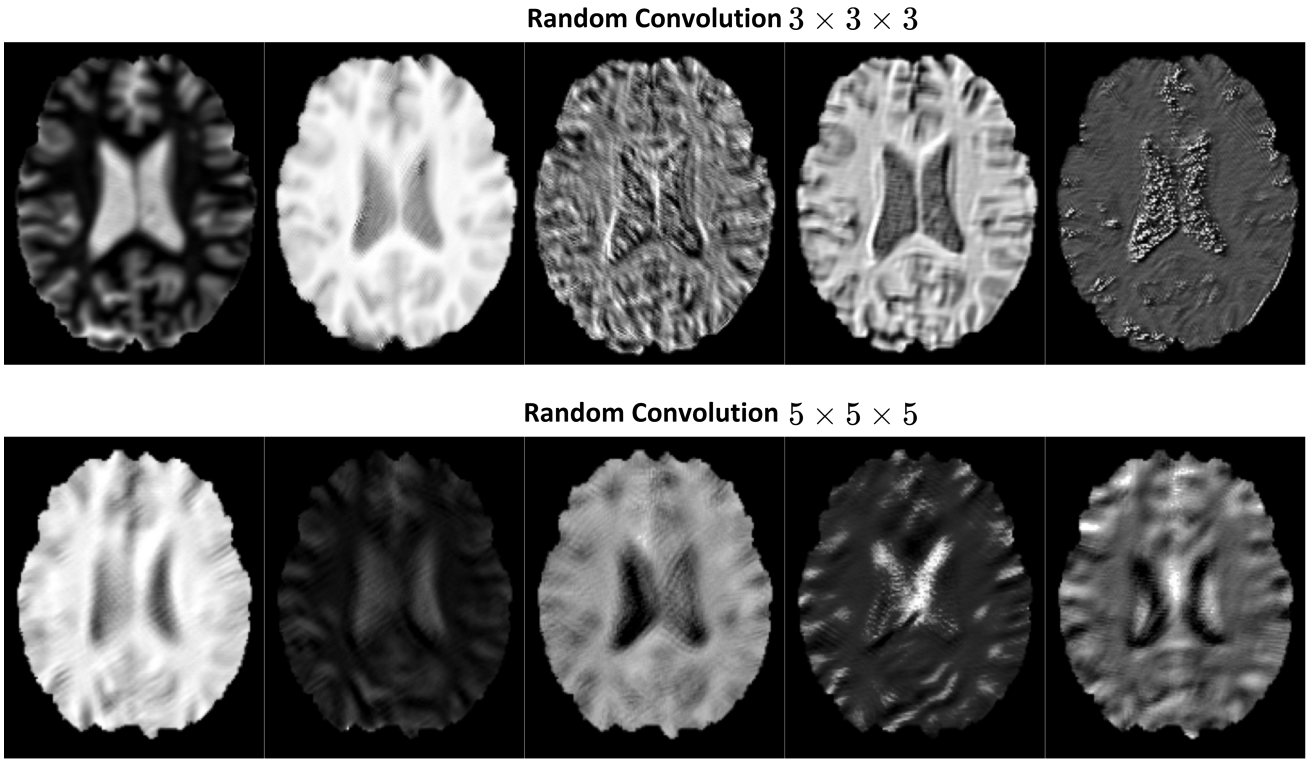

Figure S3. Axial mid-slices of augmented samples generated using the RC-based contrast augmentation method with $3 \times 3 \times 3$ and $5 \times 5 \times 5$ convolution kernels, resulting in visible blurring effects.

model ($\boldsymbol{f}(\cdot;\theta)$), the calculated deformation field (Eq. 1) is applied to deform the scans before RC augmentation (Eq. 2) and the subsequent calculation of similarity and registration loss functions. This approach is based on the fact that RC does not alter the geometric properties of the scans but instead generates arbitrary contrast variations. This forces the model to predict landmarks independently of their contrasts. Consequently, this enables the use of a mono-modal loss function in Eq. 2, such as mean square error (MSE), while eliminating the need for computationally expensive metrics like mutual information (MI), normalized mutual information (NMI), or descriptors like modality independent neighborhood descriptor (MIND).

3. Baselines

It is important to note that we did not include the 3D U-Net as one of our baselines for direct landmark detection because it failed to converge and performed poorly on the publicly available test sets. This outcome was expected, as 3D U-Net typically has a much heavier parameter load compared to simpler architectures like the 3D supervised CNN we implemented. Given our limited labeled data (122 scans), the 3D U-Net struggled to converge effectively. Therefore, we opted to use a 3D CNN as the supervised learning baseline, which is more suited for scenarios with constrained datasets.

4. Visual Comparison with ANTs and KeyMorph

Samples of landmarks generated from our proposed model, ANTs, and KeyMorph for the same subject are shown in the axial view in Fig. S4 for comparison. Note that all landmarks are in 3D. For easy visualization, we show the 3D points projected in the 2D axial view while using a mid-axial MRI slice as a reference.

5. Sensitivity to the Template Choice

To assess the sensitivity of our method to the choice of template, we tested it using the widely adopted T1-weighted Colin27 atlas [21] (a young, single-subject template) as an extreme alternative. The resulting MREs (in mm) were 5.87±4.02, 5.56±3.51, 4.92±3.12, and 5.38±3.32 for the SNSX, OASIS, HCP, and HCP-T2w datasets, respectively. While the MREs are higher than using the ICBM152 ($p<0.05$), they are on par with KeyMorph (512 KPs) with ICBM152 ($p>0.05$). This is expected, as Fonov *et al.* [17] shows that a stable group-average brain MRI template requires $\sim$160 subjects. We showed that ICBM152 didn't im-

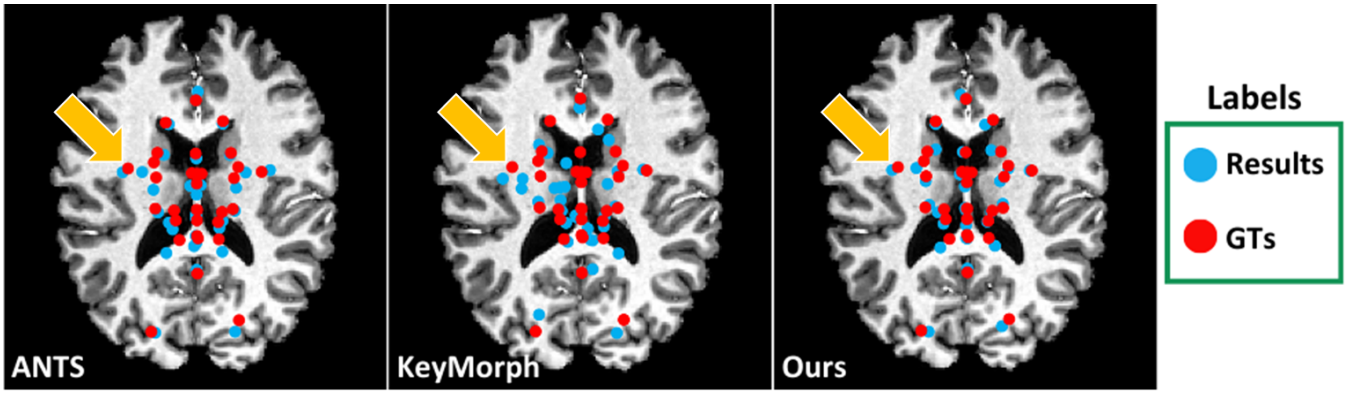

Figure S4. Qualitative comparison of anatomical landmark detection results using our model, ANTs, and KeyMorph. Note that the landmarks shown are projections of 3D points in axial view for visual demonstration.

pact the results across age groups (p>0.05). Following the suggestion of Fonov *et al*. [17], users have the flexibility to choose their own template and landmark protocol with our proposed method, as tagging a single template is not costly.

6. Robustness to Pathological Brains

While our current evaluation focuses on healthy subjects due to the availability of annotated data, assessing robustness for pathological brains is clinically important. As an indirect accuracy test, we evaluated CABLD for Parkinson's disease (PD) and Alzheimer's disease (AD) diagnosis (**Sec. 4.6**), confirming its sensitivity in detecting pathology-related landmark differences. For a direct test, we evaluated our method on 1.5T Gd-T1w MRIs of 36 PD patients from the London Health Sciences Center Parkinson's disease **(LHSCPD)** dataset [1, 53], featuring clinical scans with disease-related anatomical degeneration (e.g., atrophy) and additional domain shifts (unseen 1.5T field strength, low scan contrast/quality, and Gd MRI contrast agents). Our method achieved an **MRE=5.19±2.13mm**, not significantly different from the 40-90yo healthy group results (p>0.05), with the same age range for PD patients. Finally, upon availability of AFIDs-compliant annotations, we will validate our method for other brain disorders.

7. Computation Time

Our method achieves an average inference time of 0.35±0.012s (GPU) and 6.42±0.20s (CPU), which is significantly faster than ANTs (MI: 428.22±3.14s, CC: 380.62±0.72s, CPU), and also faster than KeyMorph (10.12±0.22s CPU, 0.54±0.01s GPU) and BrainMorph (180.14±1.99s CPU, 1.24±0.30s GPU). These demonstrate the computational efficiency of our framework for large-scale and time-sensitive applications.