



Published in final edited form as:

Neuroimage. 2020 January 01; 204: 116208. doi:10.1016/j.neuroimage.2019.116208.

Neuroimage signature from salient keypoints is highly specific to individuals and shared by close relatives

Laurent Chauvin^{a,*}, Kuldeep Kumar^a, Christian Wachinger^{e,b}, Marc Vangel^d, Jacques de Guise^a, Christian Desrosiers^a, William Wells^{c,e}, Matthew Toews^{a,**}, Alzheimer's Disease Neuroimaging Initiative

^aÉcole de Technologie Supérieure, Montreal, Canada

^bLaboratory for Artificial Intelligence in Medical Imaging, University Hospital, LMU, Munich, Germany

^cBrigham and Women's Hospital, Harvard Medical School, Boston, USA

^dMassachusetts General Hospital, Harvard Medical School, Boston, USA

^eComputer Science and Artificial Intelligence Laboratory, Massachusetts Institute of Technology, Boston, USA

Abstract

Neuroimaging studies typically adopt a common feature space for all data, which may obscure aspects of neuroanatomy only observable in subsets of a population, e.g. cortical folding patterns unique to individuals or shared by close relatives. Here, we propose to model individual variability using a distinctive keypoint signature: a set of unique, localized patterns, detected automatically in each image by a generic saliency operator. The similarity of an image pair is then quantified by the proportion of keypoints they share using a novel Jaccard-like measure of set overlap. Experiments demonstrate the keypoint method to be highly efficient and accurate, using a set of 7536 T1-weighted MRIs pooled from four public neuroimaging repositories, including twins, non-twin siblings, and 3334 unique subjects. All same-subject image pairs are identified by a similarity threshold despite confounds including aging and neurodegenerative disease progression. Outliers reveal previously unknown data labeling inconsistencies, demonstrating the usefulness of the keypoint signature as a computational tool for curating large neuroimage datasets.

Keywords

Neuroimage analysis; Individual variability; Salient image keypoints; MRI

This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

*Corresponding author. Laboratory of Imaging, Vision and Artificial Intelligence, 1100 Notre-Dame Street West, Montreal, QC H3C 1K3, Canada. laurent.chauvin0@gmail.com (L. Chauvin). **Corresponding author.: matt.toews@gmail.com (M. Toews).

Code and data availability

Complete C++ source code for the method is provided by the authors at <https://github.com/3dsift-rank/3DSIFT-Rank>. Keypoint data are available at <https://central.xnat.org/data/projects/SIFTFeatures>.

Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.neuroimage.2019.116208>.

1. Introduction

The human brain is a highly complex organ in terms of both structure and function, that is widely studied in vivo through magnetic resonance imaging (MRI) (Lerch et al., 2017). To what degree is neuroanatomy, as observed in MRI, unique to individuals? Can individuals be reliably distinguished from close relatives, i.e. siblings or monozygotic twins sharing 50–100% of their polymorphic genes, despite natural aging, neurodegenerative disease, or noise due to the data measurement process? To what degree are unique aspects of neuroanatomy shared by close relatives? These questions are motivated by increasingly personalized modern medical practices and the need to accurately curate growing sets of clinical and research neuroimaging data. We address them in this paper using a unique computer vision method.

A number of studies have investigated the variability of individuals rather than populations (Valizadeh et al., 2018; Finn et al., 2015; Miranda-Dominguez et al., 2014), where a common theme has been to encode data in terms of a unique neuroimage “fingerprint” or brainprint. Although the specific encodings used are data-dependent, the accuracy with which individuals can be identified based on their brainprint may indicate the degree to which inferences may be drawn from unique, subject-specific observations (Finn et al., 2015). Brainprint investigations have been performed from a variety of MRI modalities, including structural (Wachinger et al., 2015; Takao et al., 2015), diffusion (Valizadeh et al., 2018; Kumar et al., 2017; Yeh et al., 2016) and functional (Colclough et al., 2017; Finn et al., 2015; Miranda-Dominguez et al., 2014; Chen and Hu, 2018; Amico and Goñi, 2018) MRI, and in non-MRI data such as EEG (Armstrong et al., 2015). Our work here is the first to investigate individual identification from multiple, large-scale public MRI datasets used by the neuroimaging community, including the Human Connectome Project (HCP) (Van Essen et al., 2013), the Alzheimer’s Disease Neuroimaging Initiative (ADNI) (Jack et al., 2008) and the Open Access Series of Imaging Studies (OASIS) (Marcus et al., 2007), and first to report perfect accuracy in individual identification experiments.

The challenge in establishing a distinctive brainprint can perhaps be best illustrated by the convoluted neocortex, arguably the most distinguishing aspect of human neuroanatomy. Cortical folding patterns are highly unique to individuals and generally exhibit higher correlation between twins than unrelated individuals (Van Essen et al., 2016; Thompson et al., 2001), suggesting a link between subtle neuroanatomic structure and shared genetics. A pair-wise image correlation analysis could potentially distinguish individuals and relatives, however such an approach is generally impractical for large datasets as the number of pair-wise operations including image registration is quadratic $N(N-1)/2$ in the number of images N . Most studies of individual variability have interpreted all data in terms of a standard feature set, e.g. spatially registered voxel-wise measurements (Takao et al., 2015), neuroanatomic segmentations (Wachinger et al., 2015), cortical parcellations (Colclough et al., 2017; Finn et al., 2015; Miranda-Dominguez et al., 2014; Fischl, 2012) and related volume or thickness measurements (Sabuncu et al., 2016). While standard measurements are invaluable in interpreting data and reducing computational complexity, a number of authors have noted that a one-size-fits-all representation may obscure or average out aspects of anatomy unique to individuals (Finn et al., 2015; Gordon et al., 2017; Chen and Hu, 2018)

or close relatives, for example in the case where a one-to-one mapping between images may be ill-defined or nonexistent due to individual variability.

An alternative is to encode the image as a unique set of informative localized features or keypoints, that can be detected efficiently from generic image content and identified robustly when present in different images. For example, the highly successful scale-invariant feature transform (SIFT) (Lowe, 2004) from the field of computer vision uses highly efficient K-nearest neighbor (KNN) keypoint indexing to identify correspondences between generic intensity patterns in large image sets. Inspired by this work, we developed the 3D SIFT-Rank keypoint method (Toews and Wells, 2013) for analyzing volumetric image data as illustrated in Fig. 1. Toews et al. (Toews and Wells, 2016) showed that the proportion of detected keypoints common to a pair of brain MRIs, as quantified by the Jaccard measure of set overlap (Levandowsky and Winter 1971), could be used to identify MRI pairs of siblings with high reliability. The Jaccard measure defines a similarity matrix $\mathcal{J}(A, B)$ between all keypoints and image pairs (A, B) of a dataset, that can be used in learning-based MRI analysis. Toews et al. (Toews et al., 2010) identified class-informative keypoint clusters in the similarity matrix $\mathcal{J}(A, B)$ for MRI-based disease classification. Kumar et al. (2018) combined similarity matrices derived from multiple MRI modalities in a low-rank manifold embedding in order to study correlations between MRIs of siblings, and found that keypoints outperformed a number of baseline representations including MRI intensities and FreeSurfer derived measures (Volume + Area + Cortical Thickness). While learning procedures may be useful in analyzing a fixed dataset or group analysis, they are difficult to adapt to previously unseen data or classes, i.e. to account for images of previously unobserved individuals. This paper is the first to investigate the task of individual identification using the keypoint representation.

These results lead us to hypothesize that SIFT-Rank keypoint sets serve as a highly specific encoding of unique neuroanatomic structure. Here, we propose a novel generalisation of the Jaccard score $\mathcal{J}(A, B)$ to account for probabilistic rather than hard set equivalence between keypoints. This leads to a highly efficient instance-based inference model, allowing new MRI data to be compared to a large dataset on-the-fly in order to identify all previous scans of the same individual despite myriad potential confounds including noise, atrophy due to neurodegeneration. This paper reports the first comprehensive investigation of keypoint signatures for individual identification from MRI data. Our results based on several large-scale, public neuroimaging datasets demonstrated that all same-subject image pairs can be identified by a simple threshold on Jaccard overlap. A visual analysis of outlier cases revealed all were due to data labeling inconsistencies previously unknown to the neuroimaging community, demonstrating the practical importance of the approach in curating increasingly large data collections.

2. Material & methods

2.1. Data

Experiments are based on a large, multi-site dataset pooled from 4 public neuroimaging repositories HCP Q4, ADNI 1, OASIS 1 and OASIS 3. The FreeSurfer v6.0 pre-processing pipeline was used to remove non-brain content such as skull from MRIs while preserving

cortical and subcortical structures. Out of 8152 images, 616 failed the pre-processing pipeline (FreeSurfer error code, no output image generated) resulting in a final dataset of $N = 7536$ MRIs of 3334 unique subjects. Pre-processing failures typically occurred in Talairach alignment or in skull stripping steps, and visual inspection revealed that all cases were due to noticeable image artifacts or noise. Table 1 lists demographic and statistical information for this dataset.

Our analysis is based on a pair-wise comparison of $N(N-1)/2 = 28;391;880$ possible image pairs. Each pair is assigned a relationship label based on database metadata, for five possible sibling relationships: same subject (SM), monozygotic twins (MZ), dizygotic twins (DZ), full-sibling (FS) or unrelated subjects (UR). Subjects from different data sets are naïvely assumed to be unrelated, due to a lack of information across databases. It is important to note that these datasets were acquired under different protocols and over different periods of time, resulting in potential bias due to within-dataset similarity. For example, the time interval between scans is under a year for HCP, in comparison to 11 years for OASIS 3, 3 years for ADNI and 2 years for OASIS 1, and pairs of HCP scans may thus exhibit higher similarity than others. Our method is nevertheless robust the ranges of inter- and intra-dataset variability of these data, as we mention later in the discussion.

2.2. Processing

Assessing the pairwise similarity of images in a large data set generally requires comparing measurements at spatially homologous locations throughout the images. Since image data are noisy and the precise spatial mapping between images may be unknown or nonexistent, an effective comparison requires a combination of image registration and/or feature extraction methods. Naïve similarity assessment for all image pairs is generally intractable for large datasets as the number of pairwise operations including image registration is quadratic $N(N-1)/2$ in the number of images N , incurring a computational complexity of $O(N^2)$. Assessment based on a standard feature set such as spatially aligned cortical parcellations can reduce computational complexity, but may be insufficiently specific to capture subtle neuroanatomic patterns only observable in small subsets of a population, e.g. family members.

To address these challenges, we developed a method based on keypoint indexing (Lowe, 2004; Toews and Wells, 2013), where the image is represented as a set of generic features detected throughout the image via a saliency operator. Keypoints arise from generic neuroanatomical structure, and can be detected repeatably in a manner invariant to locally linear intensity shifts and global similarity transforms (i.e. 3D rigid transform + isotropic scaling) of image geometry (Toews and Wells, 2013). In practice, keypoint extraction is highly robust to variations in MRI intensity and geometry, e.g. in the case of images acquired from different devices or sites, exhibiting artifacts such as low frequency MRI inhomogeneity effects or changes in patient position. It is also robust to partial occlusions, where locally missing or deformed image content will not significantly impact keypoints identified in other unaffected image regions. Once detected, image content associated with keypoints is encoded into informative descriptors that can be used to identify similar keypoints via highly efficient indexing methods. Specifically, divide-and-conquer algorithms

can achieve $O(N \log N)$ complexity using search trees to identify sets of similar keypoints, thus sidestepping the need for pairwise image comparisons and scaling gracefully to large datasets. Our method consists of image keypoint detection (Fig. 1, Step I), keypoint matching (Step II) and finally computation of the Jaccard overlap similarity score (Step III). These three steps are described in greater detail below.

Keypoint Detection seeks to transform each image into a set of salient image keypoints, where keypoints are encoded as informative descriptors for efficient image content indexing. A keypoint is as a spherical image region defined by a centroid $\bar{x} = [x, y, z]$ and a scale (or size) σ within the image, and associated with a descriptor \bar{f} encoding local image appearance. A deterministic two-step detection procedure is used, including 1) salient keypoint localization and 2) keypoint encoding with the so-called SIFT-Rank approach (Toews and Wells, 2009, 2013). Keypoints are first localized by searching the image for regions that maximize an image saliency operator, signifying informative, local image patterns. A variety of such operators exist, here we use the 3D Laplacian-of-Gaussian (LoG) operator (Marr and Hildreth, 1980) that responds to generic blob-like image patterns, reminiscent of center-surround simple cell retinal processing units in the mammalian visual system (Hubel and Wiesel, 1962). The LoG operator can be approximated efficiently by the difference-of-Gaussian (DoG) operator popularized by the SIFT algorithm in 2004 (Lowe, 2004) (see Equation (1)). For each image I , a set of keypoints $\{(\bar{x}_i, \sigma_i)\}$ is identified as:

$$\{(\bar{x}_i, \sigma_i)\} = \underset{\bar{x}, \sigma}{\operatorname{argmax}} |I * G(\bar{x}, \sigma) - I * G(\bar{x}, \kappa\sigma)| \quad (1)$$

where in Equation (1) I is the 3D image, $G(\bar{x}, \sigma)$ is the Gaussian function with isotropic variance σ^2 , and κ is a constant representing the multiplicative difference in scale. Note that the keypoint scale σ is defined by the size of the Gaussian filters for which the DoG saliency operator in Equation (1) is maximized.

Once keypoint regions are localized, the image content within each region is rescaled and rotated to a characteristic coordinate system, then encoded as a descriptor \bar{f} representing the local image content as a fixed-length vector. The rescaling factor is defined by the keypoint scale σ and the rotation by the 3D orientation of local image gradients, thus the descriptor is invariant to global scaled rigid transforms (i.e. similarity transforms) of image geometry, e.g. in the cases of variable patient scanning position or unregistered images. Local image gradient analysis is also used to reject keypoints arising from image patterns that cannot be reliably localized in 3D, e.g. smooth surfaces.

Distinctive image patterns generally arise from boundaries between regions of differing intensity contrast, e.g. white and grey matter. Descriptors here encode image content as histograms of local image gradient information, estimated via finite difference operators and quantized into discrete bins over local 3D location and orientation, an approach know as the histogram-of-oriented gradient (HoG) descriptor (Lowe, 2004; Dalal and Triggs, 2005). Robustness to minor shifts or deformation of image geometry is achieved by the use of relatively coarse spatial bins, where small variations in gradient location or direction do not significantly impact descriptor values. The HoG descriptor is currently among the most

effective and widely used descriptors for image keypoint matching, and is analogous of so-called orientation hyper-columns of complex cells that encode image structure in the mammalian visual cortex (Hubel and Wiesel, 1962). In our method, local image patches are cropped and rescaled to 11^3 voxels, after which image gradient histograms are computed to encode the image content. For encoding, local 3D space and orientation are quantized uniformly into $2 \times 2 \times 2 = 8$ spatial regions and 8 orientation histogram bins, thereby producing an $8 \times 8 = 64$ element HoG image descriptor for each keypoint. Finally, this descriptor is rank-ordered to provide invariance to monotonic variations in image gradients (Toews and Wells, 2009), e.g. due to variations in image contrast.

Keypoint Matching seeks to identify pairs of keypoints arising from similar anatomical structure in different images, based on the similarity (or dissimilarity) of their descriptors. Descriptors from the same structure in different images are rarely identical, but differ by varying amounts due to imaging variations, noise, etc. The dissimilarity of a pair of descriptors (\bar{f}_i, \bar{f}_j) is quantified by the Euclidean distance or L_2 -norm between their elements $d(\bar{f}_i, \bar{f}_j) = \|\bar{f}_i - \bar{f}_j\|_2$. Assuming an additive Gaussian noise model and independent and identically distributed (IID) descriptor elements, the smaller the distance $d(\bar{f}_i, \bar{f}_j)$, the higher the likelihood that the descriptors arise from the same underlying anatomical structure.

A K-nearest neighbor (K-NN) search is used to identify sets of similar descriptors as follows. For each descriptor \bar{f}_i , a set of the K^{th} closest or most similar descriptors $NN_k(\bar{f}_i)$ is identified as

$$NN_k(\bar{f}_i) = \{\bar{f}_j : d(\bar{f}_i, \bar{f}_j) \leq d_{i,K}\} \quad (2)$$

where $d_{i,K} = d(\bar{f}_i, \bar{f}_K)$ is the distance between \bar{f}_i and the K^{th} closest descriptor \bar{f}_K . Enumerating $NN_k(\bar{f}_i)$ for each \bar{f}_i in a set of N descriptors via naïve pairwise comparisons incurs a computational cost of $\mathcal{O}(N^2)$. However rapid approximate K-NN algorithms using efficient tree-based search structures can perform this enumeration in $\mathcal{O}(N \log N)$ computational complexity. We use a set 8 randomized search trees as proposed in (Muja and Lowe, 2014), where the descriptor search space is divided according to the descriptor elements exhibiting the highest variance. Parameter K can be set generously to at least the number of expected matches in the dataset, here we used $K = 30$. Note that a relatively large percentage of matches may be spurious or incorrect due to noise, however these tend to be distributed randomly between unrelated subjects and have negligible impact. With these parameters, matching features of one image to all other 7535 images (represented by approximately 16, 000,000 features) requires approximately 0.35 s on an Intel Xeon Silver 4110@2.10Ghz, demonstrating the high computational efficiency of the method. Note that $\mathcal{O}(N^2)$ operations are required to explicitly enumerate matches between all pairs in a set of N images for a complete analysis (i.e. 28,381,485 unrelated subjects in Fig. 2), however small numbers of closely related subjects may identified in $\mathcal{O}(N \log N)$ complexity.

Set Similarity Measurement seeks to quantify the likelihood that two independent keypoint sets arise from the same underlying object. As our data consist of sets of discrete keypoints, potentially arising from equivalent neuroanatomic regions in different subject scans, we use the Jaccard overlap measure from set theory. The Jaccard overlap score $\mathcal{J}(A, B)$ between a pair of feature sets A and B is defined as

$$\mathcal{J}(A, B) = \frac{|A \cap B|}{|A \cup B|} = \frac{|A \cap B|}{|A| + |B| - |A \cap B|} \quad (3)$$

where, in Equation (3), $|A|$ and $|B|$ represent the cardinalities (sizes) of sets A and B , and $|A \cap B|$ represents the cardinality of their intersection, i.e. the of elements shared by A and B . The Jaccard measure has several desirable properties, for example it ranges from $[0,1]$ for disjoint to identical sets respectively, and can be transformed into a distance metric (Levandowsky and Winter 1971) $1 - \mathcal{J}(A, B)$ in the abstract space of variable-sized keypoints sets.

The intersection in Equation (3) is defined as the subset of elements shared by both sets:

$$A \cap B = \{\bar{f}_i: \bar{f}_i \in A \wedge \bar{f}_i \in B\} \quad (4)$$

Evaluating the intersection requires identifying pairs of equivalent set elements, i.e. $\bar{f}_i \in A$ and $\bar{f}_j \in B$ such that $\bar{f}_i = \bar{f}_j$. These are defined by the union of all unique NN keypoint matches identified between A and B . As set equivalence is binary, each unique keypoint match thus (\bar{f}_i, \bar{f}_j) contributes 1 element to the intersection $A \cap B$, and the cardinality $|A \cap B|$ may be computed as the sum of matching elements (Kumar et al., 2018; Toews and Wells, 2016). Binary equivalence of keypoint descriptors is difficult to justify however, as statistical variations in image content generally introduce uncertainty into keypoint descriptors. We thus consider (\bar{f}_i, \bar{f}_j) as contributing a soft value ranging from $[0, 1]$ to $[A \cap B]$, which is proportional to the likelihood of (\bar{f}_i, \bar{f}_j) arising from a Gaussian density of variance α_i^2 and mean \bar{f}_i . The cardinality of the set intersection is then evaluated as:

$$|A \cap B| = \sum_{\bar{f}_i \in A} \max_{\bar{f}_j \in NN_k(\bar{f}_i) \cap B} \exp\{-d^2(\bar{f}_i, \bar{f}_j)/2\alpha_i^2\} \quad (5)$$

In Equation (5), variance parameter α_i^2 is set automatically for each keypoint descriptor \bar{f}_i as the squared distance to the closest NN keypoint within the entire training set. This allows the method to adjust to variable sampling density in the keypoint descriptor space about sample \bar{f}_i in a manner similar to adaptive or variable kernel density estimation (Terrell Scott et al., 1992), thereby downweighting the contribution of unlikely matches to the cardinality $|A \cap B|$. A related benefit is the reduced sensitivity to the parameter K in the evaluation of $|A \cap B|$. Note that the cardinality for standard binary set equivalence is computed via Equation (5) by

taking $d(\bar{f}_i, \bar{f}_j) = 0$ for all $\bar{f}_j \in NN_k(\bar{f}_i)$. Given that the nearest neighbor relationship is not strictly symmetric, the cardinality as computed via Equation (5) depends generally on the feature set over which sum is computed. In practice, we notice little difference in Jaccard values, and symmetry may be imposed.

3. Results

We performed experiments to quantify the variability of T1-weighted MRI data acquired from individuals and close relatives using keypoint signatures. A large set ($N = 7536$) of T1w MRIs of 3334 unique subjects was pooled from four public neuroimaging datasets (HCP Q4, ADNI 1, OASIS 1 and OASIS 3), where each image pair bears a unique pair-wise relationship label: same subject (SM), monozygotic twin (MZ), dizygotic twin (DZ), non-twin full sibling (FS) or unrelated (UR) subjects. Relationship information for pairs of MR images were provided by individual datasets, while relationships between image pairs from different datasets were naïvely assumed to be unrelated. We evaluated the Jaccard overlap derived from the image content for all $N(N-1)/2 = 28,391,880$ possible image pairs, and studied the distributions of Jaccard measurements conditioned on relationship labels. Since the Jaccard overlap was derived from the proportion of features shared between images, we expected it to decrease with the degree of genetic and environmental separation in the pairwise relationship, i.e. decreasing in the order of SM, MZ, DZ, FS, UR pairs.

Fig. 1 illustrates the workflow for evaluating the Jaccard similarity score $\mathcal{J}(A, B)$ between an image pair (A, B) . First, a one-time pre-processing step was applied to each image, where non-brain structure was removed using the Freesurfer software (Fischl, 2012), after which keypoint features were detected using the authors' publicly available software implementation. Detection required approximately 15 s and identified approximately 2000 keypoints per MRI. After pre-processing, nearest-neighbor (NN) keypoint matches were enumerated across all images, establishing putative equivalence between keypoints in different images, and the Jaccard similarity score was computed for each image pair based on the proportions of keypoint matches they shared.

The Jaccard overlap $\mathcal{J}(A, B)$ can be viewed as a whole-brain similarity measure ranging from $[0, 1]$ for lowest to highest similarity. Equivalently, a monotonic transform such as the negative logarithm can be used to map $\mathcal{J}(A, B)$ to a Jaccard distance measure $d_{\mathcal{J}}(A, B) = -\log \mathcal{J}(A, B)$ ranging from $[0, \infty]$. Fig. 2 shows the empirical distributions of Jaccard distances, obtained for the five pair-wise relationships (indicated by color), where lower distance indicates a higher proportion of shared image content and neuroanatomic similarity. Distributions for each relationship label are unimodal and concentrated about a central tendency. The order of the mean distances for relationships (dashed vertical lines) is generally consistent with the degree of similarity in genetic and environmental developmental factors, i.e. increasing in order of SM, MZ, DZ/FS, UR. A number of outliers were identified (red and blue dots), and have been confirmed to arise from inconsistent image labels. These outliers will be discussed later.

The Jaccard distance distribution for same-subject (SM) pairs was highly unique, with no overlap with any other distribution including monozygotic twins (MZ). All SM pairs could

thus be identified via a simple score threshold, supporting our hypothesis that keypoints capture highly unique aspects of individual neuroanatomy. Distributions for other relationships exhibit a degree of overlap, and the two-sample Kolmogorov-Smirnov test was used to evaluate the null hypothesis that samples arise from the same underlying distribution (Table 2). Most p-values were extremely low, allowing us to reject the null hypothesis with high confidence. The one exception was the case of FS and DZ pairs, where the p-value of 0.108 indicated no significant difference between Jaccard distance distributions for FS and DZ siblings.

The Jaccard overlap is derived from image-to-image keypoint matches between unique neuroanatomic patterns identified in different images. Fig. 3 illustrates the spatial distributions of matching keypoints as heatmaps within a common reference space for each relationship label. Keypoints were distributed similarly throughout the brain for all relationships, and generally concentrated in regions with significant intensity contrast variations, i.e. the interfaces between cortical sulci or sub-cortical structures and cerebral spinal fluid (CSF). The primary quantitative difference between relationship groups was the number of keypoint matches, which was much higher for same subject images, thus reflecting a higher degree of shared anatomic structure. Fig. 4 shows an example of keypoints matching between MRIs of the same individual acquired 11 years apart.

The primary systematic confound was the age difference between image acquisitions, which was positively correlated with Jaccard distance as shown in Fig. 5. This likely reflected changes in brain morphology due to both natural aging and disease progression, primarily in SM pairs of older adults from the ADNI and OASIS datasets designed to study Alzheimer's disease (mean age 73 years). Data was unavailable to investigate the impact of age difference for younger SM subjects, however age separation between FS pairs (mean age 28 years) from the HCP dataset had no noticeable impact on the Jaccard distance, likely reflecting the relatively stable brain anatomy across the younger, healthy HCP cohort. By inspection, the highest Jaccard distances for SM pairs were typically associated with random confounds including image artifacts (e.g. due to MRI acquisition, pre-processing, etc.) or morphological changes in the brain (e.g. due to natural aging, neurodegenerative disease, etc.), see Fig. 5a) and b). Other confounds including sex, race and age had no significant impact on Jaccard distances, and invariant keypoint matching is independent of the image (mis)alignment, an important benefit of our method.

A surprising result was the discovery of 184 outlier image pairs with Jaccard distances that were noticeably outside of the distributions associated with their relationship labels. These included pairs labeled as UR or MZ with Jaccard distances similar to same subjects (Fig. 2, blue dots), and pairs labeled as SM with Jaccard distances similar to those for unrelated subjects (Fig. 2, red dots). Upon visual inspection of the cortical folding patterns (see examples in Fig. 2), and with the help of the respective database administrators, we established that all were most likely labeled incorrectly. All individual datasets contained at least one case where images of the same subject were labeled as different subjects. Instances of the same subject were identified across the OASIS and ADNI datasets. Table 3 lists the numbers of subjects with inconsistent labels within and between datasets. Additionally, 11

cases of perfectly identical images were identified by unusually low Jaccard distance, these were labeled as images of the same subject acquired at different time points.

An important algorithmic parameter is the number of NN descriptor matches K used in estimating the Jaccard score. In the case of hard set equivalence (Toews and Wells, 2016; Kumar et al., 2018), each keypoint match contributes 1 to the set union $|A \cup B|$, and an optimal K depends on the number of relevant image samples, i.e. MRIs of the same subject or group, which is generally unknown and variable. In the case of our soft weighting approach however, each match contributes a weight proportional to a Gaussian density, and thus K may be set large enough to include all relevant image samples.

4. Discussion

In this paper, we proposed to model neuroanatomy as a collection of distinctive image keypoints, hypothesizing that this would more accurately preserve aspects of anatomy unique to individuals or close family members, distinctive neuroanatomic signatures that might otherwise be averaged out by traditional parcellation or voxel-wise representations. The whole-brain similarity of an image pair was assessed in terms of the proportion of keypoints they share using the Jaccard measure of set overlap, which can be computed from arbitrarily large datasets using a highly efficient keypoint matching procedure.

Experiments validated our hypothesis in the largest study of individual identification from MRI data to date, involving 7536 T1w MRIs of 3334 unique subjects pooled from four large, public neuroimaging datasets: ADNI, OASIS1, OASIS3 and HCP. Distributions of Jaccard distances for same vs. unrelated subject MRI pairs are separated by a wide margin, and a simple threshold on the Jaccard distance was sufficient to identify all same-subject pairs with 100% accuracy.¹ In contrast, the largest previous study involved almost 700 subjects from the ADNI dataset alone, required multiple scans per subject as training data, and achieved less than perfect accuracy using features derived from a standard neuroanatomic segmentation (Wachinger et al., 2015).

An important potential confound is within-dataset scan similarity; same-subject scans are typically found within the same dataset, whereas scans from the same dataset are known to generally exhibit similarity due to a number of commonalities including subject demographics, age, site, scan sequence, scanner artifacts, etc (Wachinger et al., 2019). As expected, Jaccard distance distributions were generally lower within-dataset vs. across-datasets: for SM pairs (2.12 ± 0.65 vs. 4.46 ± 0.66) and for UR pairs (8.99 ± 0.76 vs: 10.01 ± 0.80), indicating a within-dataset similarity effect. However there was no overlap between distance distributions for SM and UR pairs, and we thus expect these to remain separable by a Jaccard distance threshold in new T1-weighted MRI data acquired and pre-processed (e.g. skull-stripped) with protocols similar to those used for the three datasets used here. Furthermore, the pattern of decreasing similarity in the order of SM, MZ, (DZ/FS) to UR pairs did not change when analysis was restricted to single datasets.

¹For completeness, there is the possibility of coincidental errors in both metadata labeling and automatic identification, however the probability of this is very low, given these events are unrelated, independent and individually highly improbable.

A surprising result was the discovery of previously unknown subject labeling inconsistencies, identified as clear outliers from their expected Jaccard distributions. These included MRIs of the same person labeled as unrelated, or MRIs of different people labeled as the same, and were identified both within and across the datasets. Such inconsistencies may lead to bias in cross-validation studies, e.g. computer-assisted prediction (Desikan et al., 2009; Samper-Gonzalez et al., 2018) where protocols assume independent training and testing data, and similar errors in a clinical context could potentially lead to errors in patient care. The ability to identify these in widely used, public datasets is noteworthy, and demonstrates the potential for the keypoint signature as a powerful tool in curating and validating large neuroimage datasets.

The Jaccard measure is driven by keypoint matches representing instances of unique neuroanatomic patterns shared between pairs of images of individual or siblings. The Jaccard overlap generally decreased with the degree of genetic and developmental separation in the pairwise relationship label, i.e. in the order of SM, MZ, DZ/FS, UR pairs, indicating decreasing proportions of unique anatomic structure shared by these groups as predicted. A notable exception was the case of dizygotic twin (DZ) and non-twin full sibling (FS) relationships, which showed no statistically significant difference in terms of their pairwise Jaccard distributions. Keypoint matches were generally distributed throughout the brain and across interfaces between adjacent tissues exhibiting high intensity contrast in MRI, e.g. cerebrospinal fluid, grey and white matter, in a manner unique to the specific image pair, rather than within regular loci defined by typical parcellation schemes. Combined with the high accuracy of identification experiments, this suggests that aspects of neuroanatomy most characteristic of individuals or close relatives may be highly idiosyncratic and not ultimately be tied to a fixed parcellation. Keypoint signatures provide a robust and efficient means of exploring these aspects across large datasets, and a combination of keypoint signatures and traditional segmentations or parcellations may ultimately prove most effective in understanding the variability of individuals and genetic families.

In terms of technology, the keypoint signature affords the capability of rapidly comparing a new image against a large dataset, e.g. identifying all keypoint sets with high Jaccard similarity in 0.35 s here. This is a memory-based learning approach, which requires no explicit training procedure. New data are easily incorporated, and it is limited only by the amount of memory available, a limit that is continually reduced by technological advancement. In contrast, representations based on traditional neuroanatomic parcellations generally require extensive pre-processing, including image registration and segmentation, and alternative machine learning approaches require training procedures with multiple MRIs per subject (Wachinger et al., 2015) which may be unavailable a priori. Keypoint detection here is based on a recursive Gaussian filtering process that is analogous to a highly efficient, unbiased convolutional neural network (CNN) used in deep learning (LeCun et al., 2015). Machine learning could potentially be used to optimize filters, however this would require training procedures and data that might not be readily available (i.e. multiple labeled images of an individual), and would introduce bias related to a particular training set.

Our experiments here derived keypoint signatures from the ubiquitous T1w structural MRI modality, however keypoint detection can be performed from arbitrary scalar image

modalities, e.g. fractional anisotropy derived from diffusion MRI (dMRI) (Kumar et al., 2018) or statistical parameter maps derived from fMRI data, and descriptors can be used to encode vector-valued data, e.g. histograms of diffusion gradient orientations in dMRI (Chauvin et al., 2018). Keypoint matching across different modalities is generally non-trivial and an avenue for future investigation (Toews et al., 2013). Our analysis focused on neuroanatomy and automatic skull stripping was used to remove image content associated with non-brain tissues. Extensive pre-processing is not generally required, as keypoints can be reliably detected despite variations in intensity or pose, and can be used to analyze non-brain image content. In fact, we found the Jaccard similarity between related subjects to be higher when non-brain structure is included. Nevertheless, processes for normalizing or correcting intensity values, e.g. correction of MRI inhomogeneities using field maps, should generally improve the repeatability of keypoint detection, and the question of an optimal image pre-processing pipeline is left for future research. Experiments here were limited to sibling relationships and adult subjects ranging of 18–96 years of age, future investigations will consider younger age groups such as infants with the additional confound of rapid neurodevelopment and other relationships including parent-child or cousins with varying amounts of shared genetics. Finally, our work here does not investigate links between anatomical keypoints and subject abilities or behaviors. However, the keypoint representation has previously been used to interpret single anatomical scans according to group-wise clinical symptoms or labels, e.g. Alzheimer’s disease classification (Toews et al., 2010) and neonatal age prediction (Toews et al., 2012), and similar keypoint analysis techniques could potentially be applied to other modalities including functional MRI data in future investigations.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgements

OASIS. Data were provided in part by OASIS: Cross-Sectional: Principal Investigators: D. Marcus, R. Buckner, J. Csernansky J. Morris; P50 AG05681, P01 AG03991, P01 AG026276, R01 AG021910, P20 MH071616, U24 RR021382 and OASIS-3: Principal Investigators: T. Benzinger, D. Marcus, J. Morris; NIH P50AG00561, P30NS09857781, P01AG026276, P01AG003991, R01AG043434, UL1TR000448, R01EB009352. AV-45 doses were provided by Avid Radiopharmaceuticals, a wholly owned subsidiary of Eli Lilly.

HCP. Data were provided in part by the Human Connectome Project, WU-Minn Consortium (Principal Investigators: David Van Essen and Kamil Ugurbil; 1U54MH091657) funded by the 16 NIH Institutes and Centers that support the NIH Blueprint for Neuroscience Research; and by the McDonnell Center for Systems Neuroscience at Washington University.

ADNI. Data collection and sharing for this project was funded by the Alzheimer’s Disease Neuroimaging Initiative (ADNI) (National Institutes of Health Grant U01 AG024904) and DOD ADNI (Department of Defense award number W81XWH-12-2-0012). ADNI is funded by the National Institute on Aging, the National Institute of Biomedical Imaging and Bioengineering, and through generous contributions from the following: AbbVie, Alzheimer’s Association; Alzheimer’s Drug Discovery Foundation; Araclon Biotech; BioClinica, Inc.; Biogen; Bristol-Myers Squibb Company; CereSpir, Inc.; Cogstate; Eisai Inc.; Elan Pharmaceuticals, Inc.; Eli Lilly and Company; EuroImmun; F. Hoffmann-La Roche Ltd and its affiliated company Genentech, Inc.; Fujirebio; GE Healthcare; IXICO Ltd.; Janssen Alzheimer Immunotherapy Research & Development, LLC.; Johnson & Johnson Pharmaceutical Research & Development LLC.; Lumosity; Lundbeck; Merck & Co., Inc.; Meso Scale Diagnostics, LLC.; NeuroRx Research; Neurotrack Technologies; Novartis Pharmaceuticals Corporation; Pfizer Inc.; Piramal Imaging; Servier; Takeda Pharmaceutical Company; and Transition Therapeutics. The Canadian Institutes of Health Research is providing funds to support ADNI clinical sites in Canada. Private sector contributions are facilitated by

the Foundation for the National Institutes of Health (www.fnih.org). The grantee organization is the Northern California Institute for Research and Education, and the study is coordinated by the Alzheimer's Therapeutic Research Institute at the University of Southern California. ADNI data are disseminated by the Laboratory for Neuro Imaging at the University of Southern California.

This work was supported by NIH grant P41EB015902, a Canadian National Sciences and Research Council (NSERC) Discovery Grant and the Canada Research Chair in 3D Imaging and Biomedical Engineering.

References

- Amico E, Goñi J, 2018 The quest for identifiability in human functional connectomes. *Sci. Rep* 8, 8254. [PubMed: 29844466]
- Armstrong BC, Ruiz-Blondet MV, Khalifian N, Kurtz KJ, Jin Z, Laszlo S, 2015 Brainprint: assessing the uniqueness, collectability, and permanence of a novel method for erp biometrics. *Neurocomputing* 166, 59–67.
- Chauvin L, Kumar K, Desrosiers C, De Guise J, Toews M, 2018 Diffusion orientation histograms (doh) for diffusion weighted image analysis In: *Computational Diffusion MRI*. Springer, pp. 91–99.
- Chen S, Hu X, 2018 Individual identification using the functional brain fingerprint detected by the recurrent neural network. *Brain Connect* 8, 197–204. [PubMed: 29634323]
- Colclough GL, Smith SM, Nichols TE, Winkler AM, Sotiropoulos SN, Glasser MF, Van Essen DC, Woolrich MW, 2017 The heritability of multi-modal connectivity in human brain activity. *Elife* 6, e20178. [PubMed: 28745584]
- Dalal N, Triggs B, 2005 Histograms of oriented gradients for human detection In: *International Conference on Computer Vision & Pattern Recognition (CVPR'05)*, vol.1 IEEE Computer Society, pp. 886–893.
- Desikan RS, Cabral HJ, Hess CP, Dillon WP, Glastonbury CM, Weiner MW, Schmansky NJ, Greve DN, Salat DH, Buckner RL, et al., 2009 Automated mri measures identify individuals with mild cognitive impairment and alzheimer's disease. *Brain* 132, 2048–2057. [PubMed: 19460794]
- Fedorov A, Beichel R, Kalpathy-Cramer J, Finet J, Fillion-Robin J-C, Pujol S, Bauer C, Jennings D, Fennessy F, Sonka M, et al., 2012 3d slicer as an image computing platform for the quantitative imaging network. *Magn. Reson. Imag* 30, 1323–1341.
- Finn ES, Shen X, Scheinost D, Rosenberg MD, Huang J, Chun MM, Papademetris X, Constable RT, 2015 Functional connectome fingerprinting: identifying individuals using patterns of brain connectivity. *Nat. Neurosci* 18, 1664. [PubMed: 26457551]
- Fischl B, 2012 Freesurfer, *Neuroimage* 62, 774–781. [PubMed: 22248573]
- Gordon EM, Laumann TO, Adeyemo B, Gilmore AW, Nelson SM, Dosenbach NU, Petersen SE, 2017 Individual-specific features of brain systems identified with resting state functional correlations. *Neuroimage* 146, 918–939. [PubMed: 27640749]
- Hubel DH, Wiesel TN, 1962 Receptive fields, binocular interaction and functional architecture in the cat's visual cortex. *J. Physiol* 160, 106–154. [PubMed: 14449617]
- Jack CR Jr., Bernstein MA, Fox NC, Thompson P, Alexander G, Harvey D, Borowski B, Britson PJ, Whitwell JL, Ward C, et al., 2008 The alzheimer's disease neuroimaging initiative (adni): mri methods. *J. Magn. Reson. Imaging: An Official Journal of the International Society for Magnetic Resonance in Medicine* 27, 685–691.
- Kumar K, Desrosiers C, Siddiqi K, Colliot O, Toews M, 2017 Fiberprint: a subject fingerprint based on sparse code pooling for white matter fiber analysis. *Neuroimage* 158, 242–259. [PubMed: 28684331]
- Kumar K, Toews M, Chauvin L, Colliot O, Desrosiers C, 2018 Multi-modal brain fingerprinting: a manifold approximation based framework. *Neuroimage* 183, 212–226. [PubMed: 30099077]
- LeCun Y, Bengio Y, Hinton G, 2015 Deep learning, *nature* 521, 436.
- Lerch JP, van der Kouwe AJ, Raznahan A, Paus T, Johansen-Berg H, Miller KL, Smith SM, Fischl B, Sotiropoulos SN, 2017 Studying neuroanatomy using mri. *Nat. Neurosci* 20, 314. [PubMed: 28230838]
- Levandowsky M, Winter D, 1971 Distance between sets. *Nature* 234, 34.

- Lowe DG, 2004 Distinctive image features from scale-invariant keypoints. *Int. J. Comput. Vis* 60, 91–110.
- Marcus DS, Wang TH, Parker J, Csernansky JG, Morris JC, Buckner RL, 2007 Open access series of imaging studies (oasis): cross-sectional mri data in young, middle aged, nondemented, and demented older adults. *J. Cogn. Neurosci* 19, 1498–1507. [PubMed: 17714011]
- Marr D, Hildreth E, 1980 Theory of edge detection. In: *Proceedings of the Royal Society of London. Series B. Biological Sciences*, vol. 207, pp. 187–217.
- Miranda-Dominguez O, Mills BD, Carpenter SD, Grant KA, Kroenke CD, Nigg JT, Fair DA, 2014 Connectotyping: model based fingerprinting of the functional connectome. *PLoS One* 9, e111048. [PubMed: 25386919]
- Muja M, Lowe DG, 2014 Scalable nearest neighbor algorithms for high dimensional data. *IEEE Trans. Pattern Anal. Mach. Intell* 36, 2227–2240. [PubMed: 26353063]
- Sabuncu MR, Ge T, Holmes AJ, Smoller JW, Buckner RL, Fischl B, Initiative ADN, et al., 2016 Morphometricity as a measure of the neuroanatomical signature of a trait. *Proc. Natl. Acad. Sci* 113, E5749–E5756. [PubMed: 27613854]
- Samper-Gonzalez J, Burgos N, Bottani S, Fontanella S, Lu P, Marcoux A, Routier A, Guillon J, Bacci M, Wen J, et al., 2018 Reproducible evaluation of classification methods in alzheimer’s disease: framework and application to mri and pet data. *Neuroimage* 183, 504–521. [PubMed: 30130647]
- Takao H, Hayashi N, Ohtomo K, 2015 Brain morphology is individual-specific information. *Magn. Reson. Imag* 33, 816–821.
- Terrell GR, Scott DW, et al., 1992 Variable kernel density estimation. *Ann. Stat* 20, 1236–1265.
- Thompson PM, Cannon TD, Narr KL, Van Erp T, Poutanen V-P, Huttunen M, Lönqvist J, Standertskjöld-Nordenstam C-G, Kaprio J, Khaledy M, et al., 2001 Genetic influences on brain structure. *Nat. Neurosci* 4, 1253. [PubMed: 11694885]
- Toews M, Wells W, 2009 Sift-rank: ordinal description for invariant feature correspondence In: *IEEE Conference on Computer Vision and Pattern Recognition*. IEEE, pp. 172–177.
- Toews M, Wells WM, 2013 Efficient and robust model-to-image alignment using 3d scale-invariant features. *Med. Image Anal* 17, 271–282. [PubMed: 23265799]
- Toews M, Wells WM, 2016 How are siblings similar? how similar are siblings? large-scale imaging genetics using local image features In: *IEEE 13th International Symposium on Biomedical Imaging (ISBI)*, IEEE, pp. 847–850.
- Toews M, Wells III W, Collins DL, Arbel T, 2010 Feature-based morphometry: discovering group-related anatomical patterns. *Neuroimage* 49, 2318–2327. [PubMed: 19853047]
- Toews M, Zöllei L, Wells WM, 2013 Feature-based alignment of volumetric multi-modal images In: *International Conference on Information Processing in Medical Imaging*. Springer, pp. 25–36.
- Toews M, Wells WM, Zöllei L, 2012 A feature-based developmental model of the infant brain in structural mri In: *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer, pp. 204–211.
- Valizadeh SA, Liem F, Méritat S, Hänggi J, Jäncke L, 2018 Identification of individual subjects on the basis of their brain anatomical features. *Sci. Rep* 8, 5611. [PubMed: 29618790]
- Van Essen DC, Smith SM, Barch DM, Behrens TE, Yacoub E, Ugurbil K, Consortium W-MH, et al., 2013 The Wu-minn human connectome project: an overview. *Neuroimage* 80, 62–79. [PubMed: 23684880]
- Van Essen DC, Donahue C, Dierker DL, Glasser MF, 2016 Parcellations and connectivity patterns in human and macaque cerebral cortex In: *Micro-, Meso-And Macro-Connectomics of the Brain*. Springer, Cham, pp. 89–106.
- Wachinger C, Golland P, Kremen W, Fischl B, Reuter M, Initiative ADN, et al., 2015 Brainprint: a discriminative characterization of brain morphology. *Neuroimage* 109, 232–248. [PubMed: 25613439]
- Wachinger C, Becker BG, Rieckmann A, Pölsterl S, 2019 Quantifying Confounding Bias in Neuroimaging Datasets with Causal Inference, p. 04102arXiv preprint arXiv: 1907.
- Yeh F-C, Vettel JM, Singh A, Poczos B, Grafton ST, Erickson KL, Tseng W-YI, Verstynen TD, 2016 Quantifying differences and similarities in whole-brain white matter architecture using local connectome fingerprints. *PLoS Comput. Biol* 12, e1005203. [PubMed: 27846212]

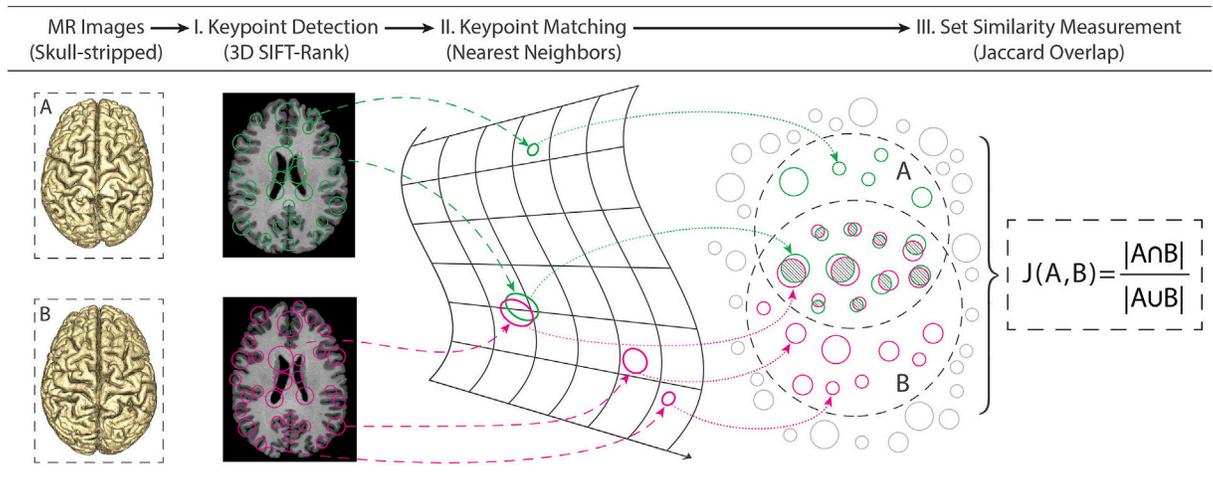


Fig. 1.

The workflow for computing the Jaccard overlap $J(A, B)$ similarity score between two images A and B. **Step I.** SIFT-Rank keypoints are extracted from skull-stripped MRI data.

Step II. Similar keypoints are identified between images using a K-nearest neighbor search.

Step III. The Jaccard overlap is computed as ratio of the intersection vs. the union of keypoint sets.

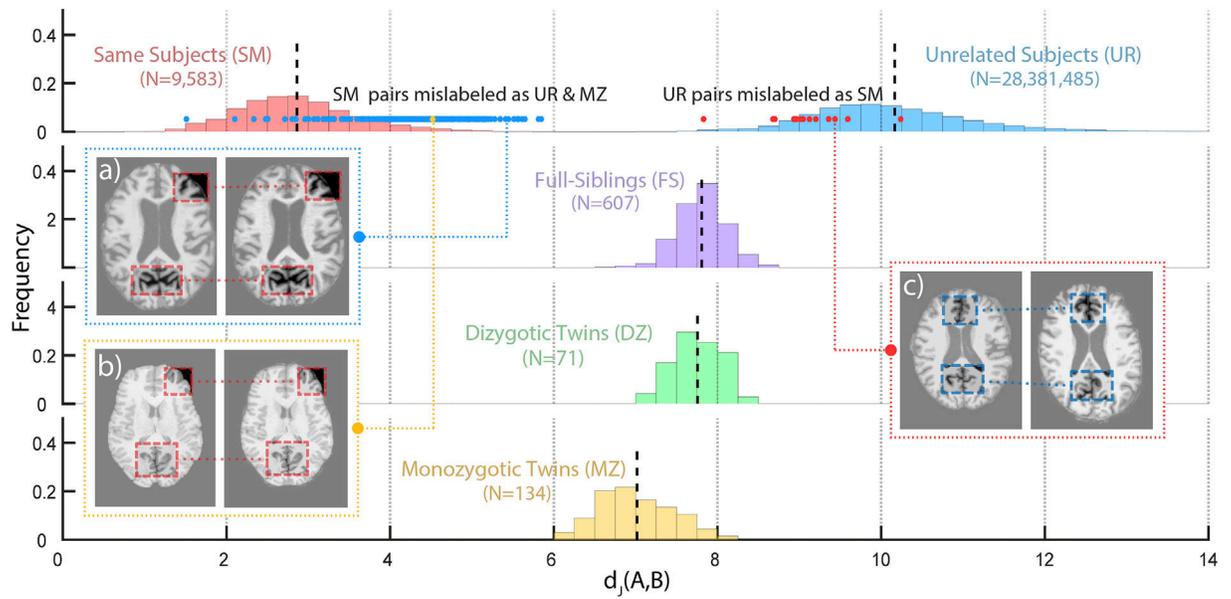


Fig. 2. Distributions of the pairwise Jaccard distances conditional on relationship labels including: unrelated (UR, blue), same (SM, red), full siblings (FS, purple), dizygotic (DZ, green) and monozygotic (MZ, yellow) twin subject relationship labels, where black dashed lines indicate distribution means. Note the high degree of separation between SM and UR scores. Dots indicate data labeling inconsistencies automatically flagged by unexpected Jaccard distance. The correct relationship labels are evident upon visual inspection of cortical patterns (highlight), which are virtually identical in SM pairs mislabeled as (a) UR and (b) MZ, and highly different in a UR pair mislabeled as (c) SM.

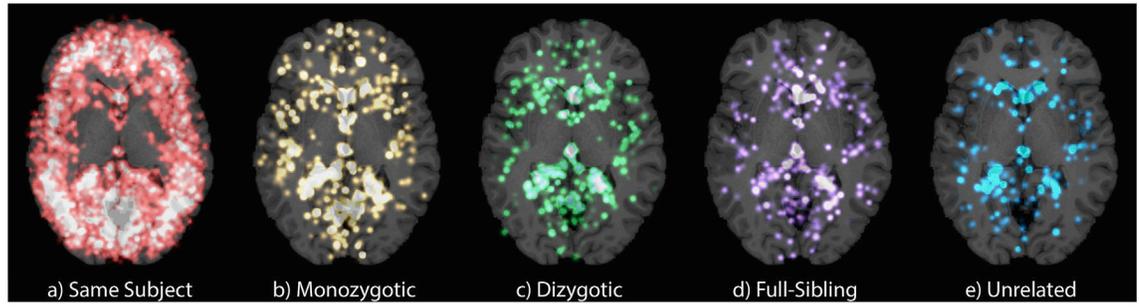


Fig. 3. The spatial layout of keypoint matches for five pairwise relationship labels: a) SM, b) MZ, c) DZ, d) FS, e) UR. Heatmaps represent distributions of matching keypoints within the standard MNI305 neuroanatomic reference space and accumulated over 71 image pairs per label from HCP dataset.

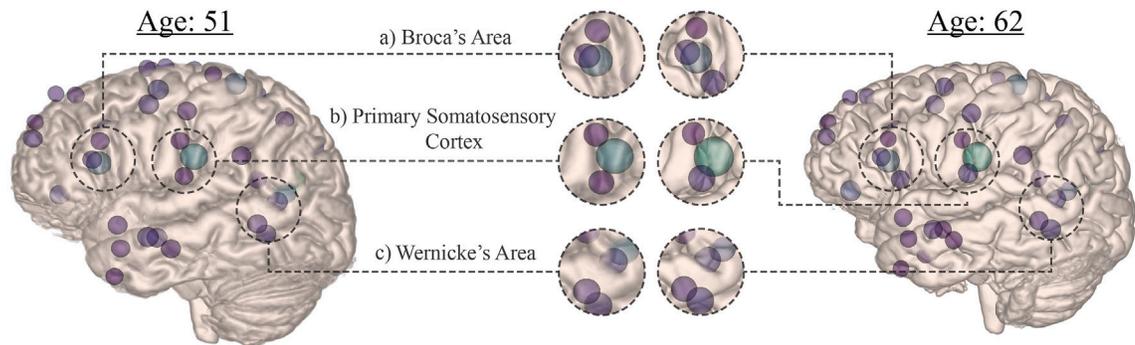


Fig. 4.

Keypoints matched between MRIs of the same individual acquired at 51 and 62 years of age. Matching keypoints (spheres, color indicating scale) represent patterns of local cortical structure that are highly unique to this individual, with notable concentrations in a) Broca's area, b) the primary somatosensory cortex and c) Wernicke's area. Note that keypoints have been slightly extruded from within the cortex for improved visualization. The visualization was generated using the 3D Slicer software (Fedorov et al., 2012).

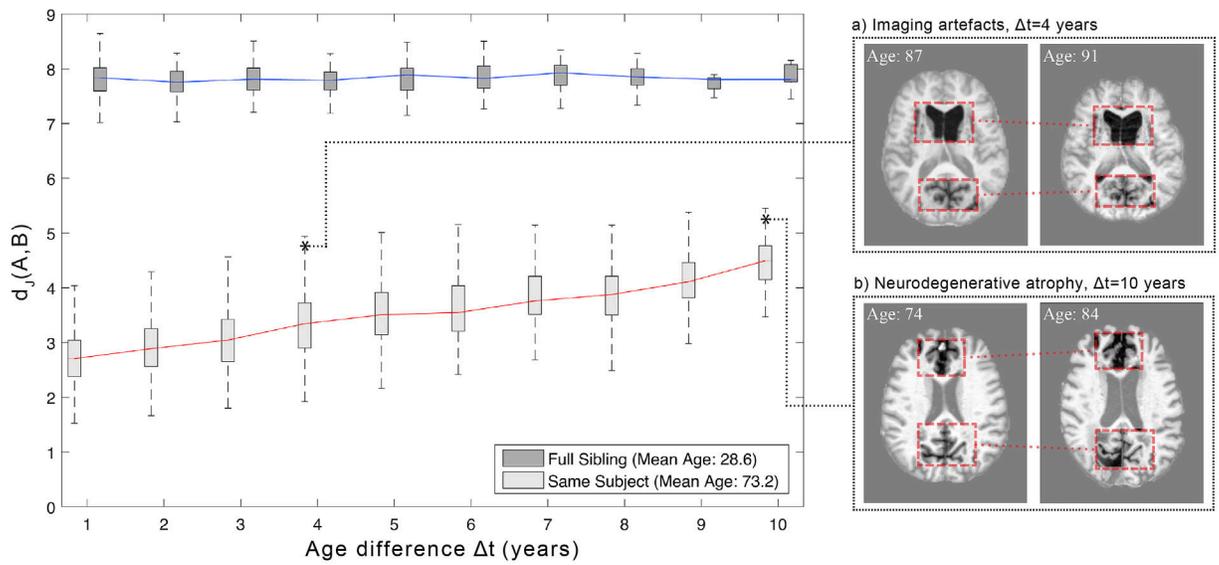


Fig. 5. Jaccard distance as a function of the age difference Δt between scans for SM ($N = 9885$) and FS ($N = 607$) pairs. Lines and boxes represent the means and standard deviations of the Jaccard distance. a) and b) show examples of image pairs successfully identified as SM, despite a relatively high Jaccard distance due to aging, noticeable image artifacts in a) and neurodegenerative atrophy in b).

Table 1

Dataset demographic and statistical information.

Dataset	Subjects	Gender (M/F)	Age (Min/Avg/Max)	Images	Voxel Size (mm)	keypoints (Avg/Image)
HCP Q4	1011	469/542	22/29/36	1053	0.7	3633
ADNI 1	844	488/356	55/75/91	3291	1.0	1879
OASIS 1	416	160/256	18/53/96	416	1.0	2143
OASIS 3	1063	470/593	42/70/97	2776	1.0	1856
Total	3334	1587/1747	18/56/97	7536	-	2130

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Table 2

p-values for two-sample Kolmogorov-Smirnov tests between Jaccard distance distributions for (SM) Same Subject, (MZ) Monozygotic, (DZ) Dizygotic, (FS) Full-Sibling and (UR) Unrelated pairwise relationships.

	SM	MZ	DZ	FS	UR
SM	-	1.38×10^{-233}	1.55×10^{-125}	0	0
MZ	-	-	1.40×10^{-41}	6.09×10^{-97}	1.18×10^{-228}
DZ	-	-	-	0.108	2.87×10^{-120}
FS	-	-	-	-	0
UR	-	-	-	-	-

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Table 3

Mislabeled subject relationships identified across and within databases. Most cases are subjects mislabeled as UR. * A pair labeled as MZ was established to be SM and one dataset was subsequently removed from the HCP dataset. A pair of scans labeled as UR were confirmed to be SM in the OASIS 1 dataset. ** OASIS3 and OASIS1 are known to share data from numerous subjects. *** OASIS3 contained 2 pairs of SM subjects mislabeled as UR, and 3 pairs of UR subjects labeled as SM.

	HCP Q4	ADNI 1	OASIS 1	OASIS 3
HCP Q4	1*	0	0	0
ADNI 1	-	3	4	2
OASIS 1	-	-	1	79**
OASIS 3	-	-	-	2 + 3***