Deep Bayesian Hashing With Center Prior for Multi-Modal Neuroimage Retrieval

Erkun Yang, Mingxia Liu, *Senior Member, IEEE*, Dongren Yao, Bing Cao, Chunfeng Lian, *Member, IEEE*, Pew-Thian Yap, *Senior Member, IEEE*, and Dinggang Shen, *Fellow, IEEE*

***Abstract*— Multi-modal neuroimage retrieval has greatly facilitated the efficiency and accuracy of decision making in clinical practice by providing physicians with previous cases (with visually similar neuroimages) and corresponding treatment records. However, existing methods for image retrieval usually fail when applied directly to multi-modal neuroimage databases, since neuroimages generally have smaller inter-class variation and larger inter-modal discrepancy compared to natural images. To this end, we propose a deep Bayesian hash learning framework, called CenterHash, which can map multi-modal data into a shared Hamming space and learn discriminative hash codes from imbalanced multi-modal neuroimages. The key idea to tackle the small inter-class variation and large inter-modal discrepancy is to learn a common center representation for similar neuroimages from different modalities and encourage hash codes to be explicitly close to their corresponding center representations. Specifically, we measure the similarity between hash codes and their corresponding center representations and treat it as a center prior in the proposed Bayesian learning framework. A weighted contrastive likelihood loss function is also developed to facilitate hash learning from imbalanced neuroimage pairs. Comprehensive empirical evidence shows that our method can generate effective hash codes and yield state-of-the-art performance in cross-modal retrieval on three multi-modal neuroimage datasets.**

***Index Terms*— Deep Bayesian hashing, retrieval, multi-modal neuroimage, MRI, PET.**

Manuscript received August 13, 2020; revised September 29, 2020; accepted October 4, 2020. Date of publication October 13, 2020; date of current version February 2, 2021. This work was supported in part by NIH under Grant AG041721 and Grant AG053867. *(Corresponding authors: Mingxia Liu; Dinggang Shen.)*

Erkun Yang, Mingxia Liu, Chunfeng Lian, and Pew-Thian Yap are with the Department of Radiology, The University of North Carolina at Chapel Hill, Chapel Hill, NC 27599 USA, and also with BRIC, The University of North Carolina at Chapel Hill, Chapel Hill, NC 27599 USA (e-mail: ekyang@med.unc.edu; mxliu@med.unc.edu; chunfeng_lian@med.unc.edu; ptyap@med.unc.edu).

Dongren Yao is with the Department of Radiology, The University of North Carolina at Chapel Hill, Chapel Hill, NC 27599 USA, also with BRIC, The University of North Carolina at Chapel Hill, Chapel Hill, NC 27599 USA, also with the Brainnetome Center, Institute of Automation, University of Chinese Academy of Sciences, Beijing 100190, China, and also with the National Laboratory of Pattern Recognition, Institute of Automation, University of Chinese Academy of Sciences, Beijing 100190, China (e-mail: yaodongren2016@ia.ac.cn).

Bing Cao is with the Department of Radiology, The University of North Carolina at Chapel Hill, Chapel Hill, NC 27599 USA, also with BRIC, The University of North Carolina at Chapel Hill, Chapel Hill, NC 27599 USA, and also with the State Key Laboratory of Integrated Services Networks, Xidian University, Xi'an 710071, China (e-mail: bcao@stu.xidian.edu.cn).

Dinggang Shen is with the Department of Radiology, The University of North Carolina at Chapel Hill, Chapel Hill, NC 27599 USA, also with BRIC, The University of North Carolina at Chapel Hill, Chapel Hill, NC 27599 USA, and also with the Department of Artificial Intelligence, Korea University, Seoul 02841, South Korea (e-mail: dinggang.shen@gmail.com).

This article has supplementary downloadable material available at https://ieeexplore.ieee.org, provided by the authors.

Color versions of one or more of the figures in this article are available online at https://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TMI.2020.3030752

I. Introduction

NEUROIMAGE analysis has made a profound contribution to modern clinical analysis [1]–[3], image-guided surgery [4]–[6], and automated diagnostic studies [7]–[13]. Currently, various digital imaging techniques have been developed to generating heterogeneous visual representations of brain tissues, such as structural magnetic resonance imaging (sMRI) [14]–[16], positron emission tomography (PET) [17], and computed tomography (CT) [18]. However, interpreting neuroimages is a complicated task, which usually requires extensive professional knowledge. In practice, it is important to provide physicians with previous cases (with visually similar neuroimages) and their corresponding treatment records to facilitate case-based reasoning and evidence-based medicine in clinical decision making [19], [20]. Hence, multi-modal neuroimage retrieval [21]–[23], which can return similar cases from heterogeneous neuroimage databases, has attracted increasing interest in the field.

In this article, we focus on hashing-based multi-modal neuroimage retrieval, which makes a good balance between the retrieval quality and computation cost. Generally, cross-modal hashing aims to map heterogeneous inputs into binary codes in a common Hamming space. Based on the availability of supervisory signals, existing methods can be roughly categorized into two groups: (1) *unsupervised hashing* [24] that learns hash functions based on original data structures and distributions; and (2) *supervised hashing* [25] that learns hash functions by exploiting original data and their semantic labels. Recently, many unsupervised and supervised cross-modal hashing methods have been developed for natural image retrieval [26]–[29]. Unfortunately, existing approaches usually achieve sub-optimal results when applied directly to multi-modal neuroimage databases for the following reasons. On the one hand, neuroimages generally contain complex tissue textures and anatomical structures compared to natural images. And subtle lesions in the local brain region can significantly affect the diagnostic results with high confidence [8], implying that

0278-0062 © 2020 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission.
See https://www.ieee.org/publications/rights/index.html for more information.

Authorized licensed use limited to: University of North Carolina at Chapel Hill. Downloaded on February 22,2021 at 16:49:06 UTC from IEEE Xplore. Restrictions apply.

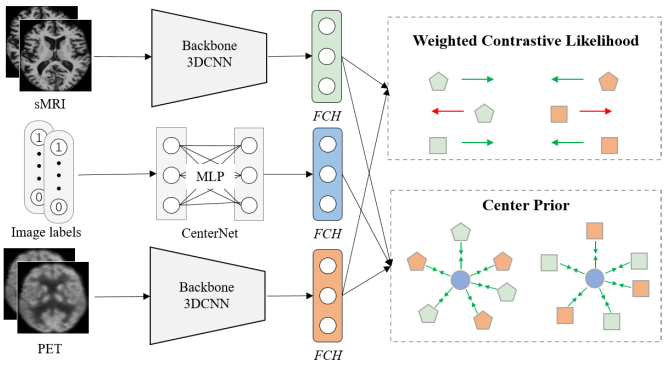


Fig. 1. Proposed deep cross-modal hashing method (called CenterHash) for multi-modal neuroimage retrieval, including three key components: (1) two modality-specific networks (*i.e.*, Backbone 3DCNNs) to project sMRI and PET to binary hash codes; (2) a CenterNet to learn center representations for each category in the Hamming space, with a multi-layer perceptron (MLP) architecture; (3) a weighted contrastive likelihood loss and a center prior, both derived from a Bayesian learning framework. Different colors denote different modalities, and different markers (except circles) denote different categories. Blue circles denote to-be-learned center representations. FCH: Fully-Connected Hash layer.

neuroimages may show *small inter-class variation*. On the other hand, even for the same subject, different neuroimaging techniques can generate different visual representations (e.g., a pair of sMRI and PET scans from the same subject), which introduces *large inter-modal discrepancy*. It's highly desired to develop advanced hashing techniques to tackle the problems of small inter-class variation and large inter-modal difference for efficient retrieval of multi-modal neuroimages.

To this end, we propose a novel deep cross-modal neuroimage hashing method (see Fig. 1), dubbed **CenterHash**, which aims to map neuroimages from different modalities into a common Hamming space and well preserve cross-modal semantic similarities. Specifically, we devise a deep multi-modal Bayesian hash learning framework, which can simultaneously learn deep representations and binary hash codes for multi-modal neuroimages. By assuming that optimal binary codes of subjects from the same class (even with different modalities) should be close to a common representation, we propose a *center prior* for the learned binary codes, which serves as the key component to address the challenges of small inter-class variation and large inter-modal discrepancy. We also propose a *weighted contrastive likelihood loss function* to enable effective hash code learning from imbalanced neuroimage pairs. Extensive experiments testify that CenterHash yields state-of-the-art performance on three multi-modal neuroimage datasets.

II. Related Work

We now briefly review the most relevant studies, including continuous representation and binary representation (via hash codes) based methods for multi/cross-modal image retrieval. For comprehensive surveys, we refer readers to [30] and [31].

A. Continuous Representation Based Methods

Conventional approaches typically rely on continuous image representations for medical image retrieval [21]–[23], [32]. For instance, Cao *et al.* [21] extended the probabilistic Latent Semantic Analysis (pLSA) model to integrate the visual and textural features of medical images and developed a deep Boltzmann machine-based multi-modal learning framework to derive the missing modality. Vikram *et al.* [22] proposed a Latent Dirichlet Allocation-based (LDA) technique to encode the visual features and explored early fusion and late fusion to combine visual features with textual features. Gao *et al.* [23] proposed a multi-graph learning-based method, which consists of two stages: query category prediction and ranking. Even though these methods have achieved good results, they usually suffer from the curse of feature dimension and are not suitable for modern neuroimage search systems.

B. Binary Representation Based Methods

To improve the efficiency of image retrieval, various fast multi-modal image retrieval methods have been proposed based on binary representation (*e.g.*, hashing codes). According to whether supervisory signals are involved in the learning phase, existing multi-modal hashing methods can be divided into two groups: (1) unsupervised approaches, and (2) supervised methods. In the following, we summarize some representative methods from these two groups, respectively.

Unsupervised approaches usually learn hash functions from the original feature space to Hamming space by exploiting the data structure, topological information, and data distributions. Semantic-topic multi-modal hashing (STMH) [33] explains multi-modal data by a composition of semantic topics, and uses binary codes to indicate the existence of the corresponding topic. Alternating co-quantization (ACQ) [34] learns hash functions by simultaneously minimizing binary quantizers for each modality and preserving data similarities across different modalities. Unsupervised generative adversarial cross-modal hashing (UGACH) [28] proposes a correlation graph to capture the manifold structure and exploits generative adversarial networks (GANs) to match the generated data pairs and pairs from the correlation graph. Unsupervised deep cross-modal hashing (UDCMH) [29] solves the hashing problem by integrating deep learning and matrix factorization.

Supervised cross-modal hashing methods [25], [35] can explore the semantic information to enhance data correlation across different modalities, thus usually achieving superior performance compared to their unsupervised counterparts. The cross-view hashing (CVH) method [25] tries to optimize the similarity-weighted cumulative Hamming distance between different pairwise samples to learn the hash functions. Semantic correlation maximization (SCM) [36] seamlessly integrates semantic labels into the hashing learning procedure and can use all the supervised information with linear-time complexity. Semantics preserving hashing (SEPH) [27] approximates the semantic affinities in Hamming space and uses the obtained hash codes as supervised signals to learn hash functions via kernel logistic regression. Though these methods have progressed the field, they typically employ hand-crafted image features, which may degrade the retrieval performance.

Recently, deep learning has revolutionized computer vision, machine learning, and many other related areas. Deep learning-based cross-modal hashing methods [26], [37]–[39]

Authorized licensed use limited to: University of North Carolina at Chapel Hill. Downloaded on February 22,2021 at 16:49:06 UTC from IEEE Xplore. Restrictions apply.

have also been proposed for image retrieval. Deep cross-modal hashing (DCMH) [37] exploits pairwise labels across different modalities to preserve the semantic relationship in Hamming space. Pairwise relationship guided deep hashing (PGDH) [38] considers both intra-modal and inter-modal similarities and designs a regularization term to maximize the representation ability of the hash codes. Ranking-based deep cross-modal hashing (RDCMH) [40] learns hash functions by utilizing a multi-level ranking semantic structure from multi-labels. Self-supervised adversarial hashing (SSAH) [39] adopts adversarial learning to maximize the representation correlations between different modalities and designs a self-supervised semantic network to discover semantic information from multi-label annotations. Cross-modal deep variational hashing (CMDVH) [41] designs cross-modal fusion networks and modality-specific networks and utilizes variational learning to match the learned hash codes. Cross-modal Hamming hashing (CMHH) [26] utilizes the exponential distribution to push relevant data pairs to have further small Hamming distances. Note that these approaches usually focus on natural images, without considering the unique property of neuroimages (*i.e.*, small inter-class variance and large inter-modal discrepancy). Thus, they may result in sub-optimal performance when applied directly to neuroimaging data.

III. Methodology

For multi-modal neuroimage search, the modality of query subjects could be different from that of subjects in the database. Here, we assume that $\mathcal{D} = \{\mathcal{X}, \mathcal{Y}\}$ is the given multi-modal neuroimage dataset, where $\mathcal{X}$ and $\mathcal{Y}$ are medical image collections from two different modalities (*e.g.*, sMRI and PET). Given o as the number of categories, $\mathcal{X} = \{\boldsymbol{x}_i\}^n$ contains n examples and $\mathcal{L}^x = \{\boldsymbol{l}_i^x\}^n$ is the label set, where $\boldsymbol{l}_i^x = [l_{i1}^x, l_{i2}^x, \cdots, l_{io}^x] \in \{0,1\}^o$ is the corresponding label for $\boldsymbol{x}_i$, $\mathcal{Y} = \{\boldsymbol{y}_i\}^m$ contains m examples and $\mathcal{L}^y = \{\boldsymbol{l}_i^y\}^n$ is the label set, where $\boldsymbol{l}_i^y = [l_{i1}^y, l_{i2}^y, \cdots, l_{io}^x] \in \{0,1\}^o$ is the corresponding label for $\boldsymbol{y}_i$. l_{id}^x and l_{id}^y are binary values that represent whether $\boldsymbol{x}_i$ and $\boldsymbol{y}_i$ belong to the d-th category. The goal of multi-modal hashing is to learn modality-specific hash functions: $h_x(\boldsymbol{x}) : \boldsymbol{x} \to \{-1,+1\}^k$ and $h_y(\boldsymbol{y}) : \boldsymbol{y} \to \{-1,+1\}^k$, which can encode original samples $\boldsymbol{x}_i$ and $\boldsymbol{y}_i$ into compact k-bit hash code $\boldsymbol{b}_i^x = h_x(\boldsymbol{x}_i) = [b_{i1}^x, b_{i2}^x, \cdots, b_{ik}^x]$ and $\boldsymbol{b}_i^y = h_y(\boldsymbol{y}_i) = [b_{i1}^y, b_{i2}^y, \cdots, b_{ik}^y]$ in a common Hamming space such that the original semantic similarity can be maximally preserved, where b_{id}^x and b_{id}^y indicate the d-th element of $\boldsymbol{b}_i^x$ and $\boldsymbol{b}_i^y$ respectively.

To address the challenges of small inter-class variation and large inter-modal discrepancy and enable learning from imbalanced neuroimaging data, this article presents a deep Bayesian hash learning method (called **CenterHash**) for multi-modal neuroimage search, with the architecture shown in Fig. 1. The proposed CenterHash employs multi-modal neuroimages (*i.e.*, sMRI and PET) and their labels as input and can simultaneously learn neuroimage representations and binary hash codes through an end-to-end pipeline. Three components are included: (1) two modality-specific networks (*i.e.*, Backbone 3DCNNs) to project input images to binary hash codes; (2) the proposed CenterNet to learn center representations in Hamming space; (3) a weighted contrastive likelihood loss and a novel center prior, both derived from a Bayesian learning framework. Details can be found in the following.

A. Framework

In the customary settings for supervised hashing [37], [38], pairwise labels are usually used as supervised signals. Here, we construct a similar pairwise sample set $\mathcal{W}$, with each element denoting a pair of sMRI and PET scans from the same category. Also, we construct a dissimilar pairwise sample set $\mathcal{C}$, with each element denoting a pair of sMRI and PET scans from different categories. Accordingly, the pairwise similarity label can be defined as

$$s_{ij} = \begin{cases} 1, & (\boldsymbol{x}_i, \boldsymbol{y}_j) \in \mathcal{W}, \\ 0, & (\boldsymbol{x}_i, \boldsymbol{y}_j) \in \mathcal{C}. \end{cases} \tag{1}$$

For effective multi-modal neuroimage hashing, we want that semantically similar neuroimage pairs can have similar hash codes in Hamming space with a high probability, and vice versa. To effectively estimate this probability, we implement a deep Bayesian learning framework. Specifically, given the pairwise similarity labels $\boldsymbol{S} = \{s_{ij}\}$, the maximum a posterior estimation of the learned binary hash codes $\boldsymbol{B}^x = [\boldsymbol{b}_i^x, \cdots, \boldsymbol{b}_n^x]$ for sMRI and $\boldsymbol{B}^y = [\boldsymbol{b}_i^y, \cdots, \boldsymbol{b}_m^y]$ for PET can be defined as

$$\log p(\boldsymbol{B}^x, \boldsymbol{B}^y | \boldsymbol{S}) \propto \log p(\boldsymbol{S} | \boldsymbol{B}^x, \boldsymbol{B}^y) p(\boldsymbol{B}^x) p(\boldsymbol{B}^y), \tag{2}$$

where $p(\boldsymbol{B}^x)$ and $p(\boldsymbol{B}^y)$ are prior distributions for the learned hash codes, $p(\boldsymbol{S} | \boldsymbol{B}^x, \boldsymbol{B}^y)$ is the weighted contrastive likelihood function defined as

$$\log p(\boldsymbol{S} | \boldsymbol{B}^x, \boldsymbol{B}^y) = \sum_{s_{ij} \in \boldsymbol{S}} w_{ij} \log p(s_{ij} | \boldsymbol{b}_i^x, \boldsymbol{b}_j^y), \tag{3}$$

where w_{ij} is the corresponding weight that is used to deal with the data imbalance problem by re-weighting each data pair $(\boldsymbol{x}_i, \boldsymbol{y}_j)$ according to the importance of each category. We set w_{ij} as

$$w_{ij} = \begin{cases} 1 + \dfrac{1}{\delta}, & s_{ij} = 1, \\ 1 + \delta, & s_{ij} = 0, \end{cases} \tag{4}$$

where $\delta = |\boldsymbol{S}_1| / |\boldsymbol{S}_0|$ denotes the data imbalance rate with $\boldsymbol{S}_1 = \{s_{ij} \in \boldsymbol{S} : s_{ij} = 1\}$ and $\boldsymbol{S}_0 = \{s_{ij} \in \boldsymbol{S} : s_{ij} = 0\}$ represents the similar and dissimilar pairwise label sets respectively. $|\cdot|$ is used to indicate the cardinality of a set. $p(s_{ij} | \boldsymbol{b}_i^x, \boldsymbol{b}_j^y)$ is the conditional probability of pairwise label s_{ij} given the hash codes $\boldsymbol{b}_i^x$ and $\boldsymbol{b}_j^y$, which can be naturally defined by the Bernoulli distribution

$$\begin{aligned} p(s_{ij} | \boldsymbol{b}_i^x, \boldsymbol{b}_j^y) &= \begin{cases} \sigma(I(\boldsymbol{b}_i^x, \boldsymbol{b}_j^y)), & s_{ij} = 1, \\ 1 - \sigma(I(\boldsymbol{b}_i^x, \boldsymbol{b}_j^y)), & s_{ij} = 0, \end{cases} \\ &= \sigma(I(\boldsymbol{b}_i^x, \boldsymbol{b}_j^y))^{s_{ij}} (1 - \sigma(I(\boldsymbol{b}_i^x, \boldsymbol{b}_j^y)))^{(1 - s_{ij})}, \end{aligned} \tag{5}$$

where $\sigma(x) = 1/(1 + e^{-\alpha x})$ is the adaptive sigmoid function with α to control the bandwidth, and $I(\boldsymbol{b}_i^x, \boldsymbol{b}_j^y)$ indicates

Authorized licensed use limited to: University of North Carolina at Chapel Hill. Downloaded on February 22,2021 at 16:49:06 UTC from IEEE Xplore. Restrictions apply.

the similarity of hash codes $\boldsymbol{b}_i^x$ and $\boldsymbol{b}_j^y$ in Hamming space. Since we want to minimize the Hamming distance between similar neuroimage pairs and maximize the Hamming distance between dissimilar neuroimage pairs, we can naturally set $I(\boldsymbol{b}_i^x, \boldsymbol{b}_j^y)$ as the negative Hamming distance:

$$I(\boldsymbol{b}_i^x, \boldsymbol{b}_j^y) = -\text{dist}_H(\boldsymbol{b}_i^x, \boldsymbol{b}_j^y), \tag{6}$$

where $\text{dist}_H(\boldsymbol{b}_i^x, \boldsymbol{b}_j^y) = |b_{id}^x \neq b_{jd}^y, 1 \leq d \leq k|$ is the Hamming distance function. However, calculating $\text{dist}_H(\boldsymbol{b}_i^x, \boldsymbol{b}_j^y)$ requires discrete operations and is non-differentiable, directly optimizing Eq. (2) with back propagation is not feasible. As indicated in [42], for a pair of hash codes $\boldsymbol{b}_i^x$ and $\boldsymbol{b}_j^y$, there exists a nice relationship between their inner product $\langle \cdot, \cdot \rangle$ and Hamming distance $\text{dist}_H(\cdot, \cdot)$:

$$\text{dist}_H(\boldsymbol{b}_i^x, \boldsymbol{b}_j^y) = \frac{1}{2}(k - \langle \boldsymbol{b}_i^x, \boldsymbol{b}_j^y \rangle). \tag{7}$$

Hence, we can adopt the inner product to quantify the pairwise similarity of hash codes in Hamming space by setting $I(\boldsymbol{b}_i^x, \boldsymbol{b}_j^y) = \langle \boldsymbol{b}_i^x, \boldsymbol{b}_j^y \rangle$. Similar to binary logistic regression, we encourage that, if the hash codes $\boldsymbol{b}_i^x$ and $\boldsymbol{b}_j^y$ are similar, the conditional probability $p(s_{ij} = 1 | \boldsymbol{b}_i^x, \boldsymbol{b}_j^y)$ should be large, implying that the multi-modal image pair $\boldsymbol{x}_i$ and $\boldsymbol{y}_j$ should be semantically similar. Otherwise, they should be semantically dissimilar. Therefore, Eq. (5) can be regarded as a reasonable extension of the binary logistic regression to the pairwise Hamming classification scheme.

B. Formulation of CenterHash

With the framework proposed in Section III-A, we can easily instantiate a specific cross-modal hash model with any valid prior distribution for the learned hash codes. Many state-of-the-art hash methods neglect the prior distribution or only assume a uniform prior, and learn hash codes by optimizing the likelihood function in Eq. (5). However, directly applying these methods may fail to obtain optimal hash codes due to the unique property of multi-modal neuroimages (*i.e.*, small inter-class variation and large inter-modal discrepancy). Since hash codes for scans from the same class are expected to be similar to each other, which is consistent with clustering, where similar data points are expected to be in the same group, inspired by k-means clustering [43], here we assume that multi-modal neuroimages in the same category share a common center representation in the Hamming space. We design a novel *center prior* distribution for the learned hash codes as follows

$$\begin{aligned} p(\boldsymbol{b}_i^x) &= \sigma(I(\boldsymbol{b}_i^x, \boldsymbol{c}_{\boldsymbol{l}_i^x})), \\ p(\boldsymbol{b}_j^y) &= \sigma(I(\boldsymbol{b}_j^y, \boldsymbol{c}_{\boldsymbol{l}_j^y})), \end{aligned} \tag{8}$$

where $\boldsymbol{c}_{\boldsymbol{l}_i^x}$ and $\boldsymbol{c}_{\boldsymbol{l}_i^y}$ are the binary center representations in Hamming space for the class $\boldsymbol{l}_i^x$ and the class $\boldsymbol{l}_i^y$, respectively. Such class-specific center presentations are shared across different modalities. This prior distribution assumes that the optimal muti-modal hash codes should distribute closely to their corresponding class centers, which is consistent with the requirement that objects with same semantic labels should have similar hash codes. By maximizing the center prior in Eq. (8), one can encourage to-be-learned hash codes (of each class) from different modalities to be close to their corresponding shared class centers, thus explicitly and simultaneously enhancing the intra-class compactness and minimizing the inter-modal discrepancy.

For Eq. (8), we need to specify the definition of the center representation $\boldsymbol{c}_{\boldsymbol{l}}$. The most straightforward way is to set $\boldsymbol{c}_{\boldsymbol{l}}$ as the mean value of hash codes of two modalities (*i.e.*, $\boldsymbol{b}_i^x$ and $\boldsymbol{b}_j^y$) for the class $\boldsymbol{l}$. However, the mean operation will destroy the binary property and will also make $\boldsymbol{c}_{\boldsymbol{l}}$ suffer from the instability caused by mini-batch training. We alternatively propose a neural network (named **CenterNet**) to automatically learn the binary center presentation for each category as

$$\boldsymbol{c}_{\boldsymbol{l}} = h_l(\boldsymbol{l}), \tag{9}$$

where h_l represents the mapping function of CenterNet, which maps the label $\boldsymbol{l}$ to the binary center representation. In this way, we can learn the multi-modal hash codes and the shared class center representations simultaneously.

By taking Eqs. (5)-(6) into the maximum a posterior estimation in Eq. (2), we obtain the overall optimization objective for the proposed CenterHash as

$$\min_{\Theta} L + T, \tag{10}$$

where L is the weighted contrastive likelihood loss, T is the center prior loss, and Θ is the network parameters. Specifically, L can be formulated as

$$L = \sum_{s_{ij} \in \boldsymbol{S}} w_{ij}(\log(1 + exp(\alpha \langle \boldsymbol{b}_i^x, \boldsymbol{b}_j^y \rangle)) - \alpha s_{ij} \langle \boldsymbol{b}_i^x, \boldsymbol{b}_j^y \rangle). \tag{11}$$

Similarly, the center prior loss can be formulated as

$$\begin{aligned} T = &\sum_{i=1}^{n} \log(1 + exp(\alpha \langle \boldsymbol{b}_i^x, \boldsymbol{c}_{\boldsymbol{l}_i^x} \rangle)) - \alpha \langle \boldsymbol{b}_i^x, \boldsymbol{c}_{\boldsymbol{l}_i^x} \rangle \\ &+ \sum_{j=1}^{m} \log(1 + exp(\alpha \langle \boldsymbol{b}_j^y, \boldsymbol{c}_{\boldsymbol{l}_j^y} \rangle)) - \alpha \langle \boldsymbol{b}_j^y, \boldsymbol{c}_{\boldsymbol{l}_j^y} \rangle, \end{aligned} \tag{12}$$

where the hash codes $\boldsymbol{b}_i^x$ and $\boldsymbol{b}_j^y$ are learned from the corresponding modality-specific networks (*i.e.*, Backbone 3DCNNs in Fig. 1), and the center representations $\boldsymbol{c}_{\boldsymbol{l}_i^x}$ and $\boldsymbol{c}_{\boldsymbol{l}_j^y}$ are learned from the CenterNet. Detailed configurations for all these networks will be elaborated in Subsection III-D.

By optimizing the overall objective in Eq. (8), we can jointly learn two hash functions $h_x(\cdot)$ and $h_y(\cdot)$ and one class-specific center representation learning function $h_l(\cdot)$. Given a query neuroimage, we can obtain its hash codes by first forward propagating it from the corresponding modality-specific hash function and then thresholding it via the sign function $sgn(\cdot)$, where $sgn(h_i) = 1$ if $h_i > 0$, otherwise $sgn(h_i) = -1$. Even though the proposed CenterHash is designed to retrieve neuroimages with two modalities (*i.e.*, sMRI and PET), one can easily extend it to applications with more modalities.

Authorized licensed use limited to: University of North Carolina at Chapel Hill. Downloaded on February 22,2021 at 16:49:06 UTC from IEEE Xplore. Restrictions apply.

C. Connection to Classification

In this subsection, we reveal that there exists a close relationship between the proposed prior distribution and classification. To enhance the discriminability of the learned hash codes, many studies assume that the hash codes should be optimal for a jointly learned classifier and usually model the relationship between hash codes (e.g., $\boldsymbol{b}_i^x$) and semantic labels (e.g., $\boldsymbol{l}_i^x$) with a linear classifier $f^x(\cdot)$ defined as

$$f(\boldsymbol{b}_i^x) = \boldsymbol{W}^{x\top}\boldsymbol{b}_i^x, \tag{13}$$

where $\boldsymbol{W}^x = [\boldsymbol{w}_1^x, \boldsymbol{w}_2^x, \cdots, \boldsymbol{w}_o^x]$ is the parameter for the classifier $f^x(\cdot)$, and $\boldsymbol{w}_d^x$ is the weight vector for the d-th class. The sigmoid cross-entropy loss is usually adopted to optimize the classifier

$$\begin{aligned}\mathcal{L}_c^x(\boldsymbol{b}_i^x) &= -[\boldsymbol{l}_i^{x\top}\log\frac{1}{1+exp(-f(\boldsymbol{b}_i^x))} \\ &\quad + (1-\boldsymbol{l}_i^x)^\top \log(1-\frac{1}{1+exp(-f(\boldsymbol{b}_i^x))})] \\ &= \sum_{d=1}^{o} -[l_{id}^x \log\frac{1}{1+exp(-\boldsymbol{w}_d^{x\top}\boldsymbol{b}_i^x)} \\ &\quad + (1-l_{id}^x)\log(1-\frac{1}{1+exp(-\boldsymbol{w}_d^{x\top}\boldsymbol{b}_i^x)})]. \end{aligned}\tag{14}$$

From Eq. (14), we can observe that for all $d \in [1, 2, \cdots, o]$ and $i \in [1, 2, \cdots, n]$, if $l_{id}^x = 1$, $\boldsymbol{w}_d^{x\top}\boldsymbol{b}_i^x$ will be maximized; and it will be minimized, otherwise. This means that optimizing the classification loss in Eq. (14) will encourage all hash codes to be close to the corresponding class weight vectors and be away from weight vectors from other classes.

Considering the prior distribution in Section III-B, the class weight vectors act similar to the class centers in Eq. (8). However, the class weight vectors contain continuous values and may have different scales, while the class centers are all binary hash codes. To better compare these two vectors, we set $\boldsymbol{w}_d^x = \boldsymbol{E}_d\boldsymbol{c}_d$, where $\boldsymbol{c}_d$ is a binary vector for the d-th class and $\boldsymbol{E}_d = diag(e_1^d, e_2^d, \cdots, e_k^d)$ is a diagonal transformation matrix with all non-negative diagonal entries (*i.e.*, $\forall i \in [0, 1, \cdots, k] : e_i^d \geq 0$). By substituting $\boldsymbol{w}_d^x$ into Eq. (14), we can get

$$\begin{aligned}\mathcal{L}_c^x(\boldsymbol{b}_i^x) &= \sum_{d=1}^{o} -[l_{id}^x \log\frac{1}{1+exp(-\boldsymbol{c}_d^\top\boldsymbol{E}_d\boldsymbol{b}_i^x)} \\ &\quad + (1-l_{id}^x)\log(1-\frac{1}{1+exp(-\boldsymbol{c}_d^\top\boldsymbol{E}_d\boldsymbol{b}_i^x)})]. \end{aligned}\tag{15}$$

From Eq. (15), we can see that optimizing the loss function actually minimizes or maximizes $\boldsymbol{c}_d^\top\boldsymbol{E}_d\boldsymbol{b}_i^x$. Note that $\boldsymbol{E}_d\boldsymbol{b}_i^x = [e_1^d b_{i1}^x, e_2^d b_{i2}^x, \cdots, e_k^d b_{ik}^x]$, compared with the inner product $\boldsymbol{c}_d^\top\boldsymbol{b}_i^x$, $\boldsymbol{c}_d^\top\boldsymbol{E}_d\boldsymbol{b}_i^x$ re-weights the i-th hash dimension with e_i^d for the d-th class. Since larger weights indicate more important hash dimensions, learning with Eq. (15) will make the learned hash codes focus on some specific hash dimensions with large weights. In addition, since the weight vectors for different classes are different, the hash dimension weights may be inconsistent across different classes, which can greatly hinder the accurate Hamming distance calculation and degrade the search performance. Furthermore, different scales of the weight vectors for different classes will make the learned hash codes focus on some specific classes and hinder the model to generalize to other classes. Thus learning with the linear classifier in Eq. (14) may be not optimal for hash code learning. To this end, our method explicitly learns the binary class centers $\boldsymbol{c}_{l_d}$ and directly minimizes the distances between hash codes and their class centers, thus can generate more consistent hash codes with better generalization ability.

TABLE I
THE CONFIGURATION OF CENTERNET. FC: FULLY-CONNECTED

Layer	Number of Units
FC-1	(label dimension o + length of hash code k)/2
FC-2	length of hash code k

D. Implementation

The architecture of modality-specific networks for sMRI and PET (*i.e.*, $\mathcal{X}$ and $\mathcal{Y}$) are similar to the 3DCNN model designed in [16], with the last fully-connected (FC) layer replaced by a new FC layer with k units as the hash layer. Besides, $\tanh(\cdot)$ is used as the activation function to relaxed the binary codes. In the *Supplementary Materials*, we show that optimizing both of the loss functions L and T with the relaxed hash codes obtained from $\tanh(\cdot)$ will decrease the loss values with exact binary codes in each iteration. For our CenterNet, which is designed to learn class-specific center representations, we use a multi-layer perceptron (MLP) that consists of two FC layers, denoted as "FC-1" and "FC-2". Besides, ReLU is used as the activation function for FC-1, while the $\tanh(\cdot)$ function is employed for FC-2. The detailed configuration of CenterNet is shown in Table I. The learning algorithm, optimization procedure, and the computation complexity analysis can be found in the *Supplementary Materials*. In the experiments, we randomly sample data points from the training set to form a mini-batch. Then, for each iteration, the similar and dissimilar data pairs are constructed based on all data points in the mini-batch.

IV. EXPERIMENTS

A. Datasets

We evaluate our method on three popular benchmark datasets: (1) Alzheimer's Disease Neuroimaging Initiative (**ADNI1**), (2) **ANDI2** [44], and (3) Australian Imaging, Biomarkers and Lifestyle dataset (**AIBL**) [45]. In the following, we give a more detailed introduction for each dataset.

ADNI1 contains 821 subjects with 1.5T T1-weighted sMRI scans, and only 397 subjects among them have PET images. Each subject was annotated by a category-level label, *i.e.*, Alzheimer's disease (AD), normal control (NC), or mild cognitive impairment (MCI). Such labels were determined based on the standard clinical criteria, including mini-mental state examination scores and clinical dementia rating. Among subjects with sMRI scans in ADNI1, there are 229 NC, 393 MCI, and 199 AD subjects. For PET data in ADNI1, there are 100 NC, 93 AD, and 204 MCI subjects.

ADNI2 contains 636 subjects with 3T T1-weighted sMRI scans, and 309 subjects have PET images. With the same

Authorized licensed use limited to: University of North Carolina at Chapel Hill. Downloaded on February 22,2021 at 16:49:06 UTC from IEEE Xplore. Restrictions apply.

clinical criteria for ADNI1, these images were divided into three categories (*i.e.*, AD, NC, and MCI). For sMRI data, there have 200 NC, 277 MCI, and 159 AD subjects. For PET data, there have 94 NC, 149 MCI, and 66 AD subjects.

AIBL consists of 612 subjects with 3T T1-weighted sMRI scans and 560 subjects with PET scans. Similar to ADNI1 and ADNI2, these images are grouped into AD, NC, and MCI, respectively. For sMRI modality, there are 447 NC, 94 MCI, and 71 AD subjects. For PET modality, there are 407 NC, 91 MCI, and 62 AD subjects.

For all the three datasets, we randomly select 10% images from each class to form the test set, and the remaining images are used as the retrieval set. To optimize the proposed method, we further randomly select 90% images from the retrieval set as the training set and treat the others as the validation set. Following [16], we pre-process all sMRI and PET scans using a standard pipeline, including anterior commissure (AC)-posterior commissure (PC) correction, intensity correction [46], skull stripping [47], and cerebellum removing. The affine registration is also performed to align each PET image to its corresponding sMRI scan.

B. Competing Methods and Evaluation Metrics

The proposed CenterHash method is first compared with five state-of-the-art cross-modal hashing methods based on traditional machine learning techniques, including ACQ [34], STMH [33], CVH [25], SEPH [27], and SCM [36] and then compared with three recently proposed deep learning methods for cross-modal hashing, including PGDH [38], DCMH [37], and CMHH [26]. The codes of the first seven competing methods are kindly provided by the authors. For CMHH, we implement the method by ourselves. Five traditional methods (*i.e.*, ACQ, STMH, CVH, SCM, and SEPH) are implemented with MATLAB, while four deep learning methods (*i.e.*, DCMH, PGDH, CMHH, and our CenterHash) are implemented with Keras [48] or Tensorflow [49]. Among eight competing approaches, two methods (*i.e.*, ACQ and STMH) are unsupervised, while the remaining ones are supervised.

For traditional methods, we extract volumes of the grey matter from 90 regions-of-interest (ROIs) as features for representing sMR and PET images, and use these ROI features as input. For deep learning methods, we use raw images as inputs. All parameters from the modality-specific 3DCNNs and CenterNet in our CenterHash are randomly initialized. To accelerate the model training, we first pre-train each modality-specific 3DCNN by simply optimizing a single-modal version of Eq. 5. Then, we combine the modality-specific 3DCNNs and CenterNet together to simultaneously optimize the hash codes for different modalities and the center representations. For a fair comparison, DCMH and PGDH use the same backbone network architectures and pre-training strategy as our method. Adam is used to optimize the network parameters of four deep learning methods (*i.e.*, DCMH, PGDH, CMHH, and CenterHash), where the initial learning rate is set as 0.01 and the mini-batch size is fixed as 5. Similar to [50], [51], we select the hyper-parameter α via cross-validation.

Four evaluation metrics are used in this article, including mean of average precision (MAP), recall@K, topN-precision, and precision-recall. The first three metrics are based on Hamming space ranking, which rank returned data points based on their Hamming distances to the queries. The precision-recall metric is based on hash lookup, as it first builds a hash lookup table and returns data points within a pre-defined Hamming radius. Given a query and a list of Q ranked retrieval examples, the average precision (AP) for the query is defined as

$$AP(\boldsymbol{x}_q) = \frac{1}{R}\sum_{q=1}^{Q} P(q)\delta(q), \tag{16}$$

where R denotes the number of ground-truth relevant points in the database, and $P(q)$ denotes the precision value for the top q retrieved points. $\delta(q) = 1$ when the q-th retrieval point is relevant with the query, otherwise $\delta(q) = 0$. In our experiments, Q is set as 100 for all the three datasets. The **MAP** is defined as the average APs for all queries. **Recall@K** is defined as the percentage of ground-truth neighbors from the top K returned instances among all the semantically relevant points. **TopN-precision** indicates the average ratio of semantically relevant instances among the top N returned points for all given queries based on Hamming distance. In this article, we set both K and N as 100. **Precision-recall** reveals the retrieval precision with different recall values and is considered as a good indicator of overall search performance.

C. Results

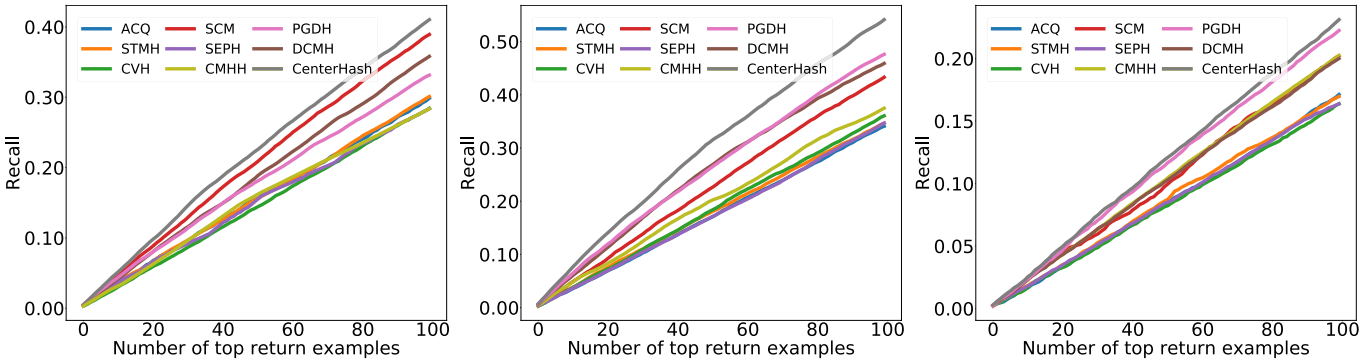
1) Results of Hamming Ranking: We first present MAP results for CenterHash and all other baseline methods on the three datasets using different lengths of hash code (*i.e.*, k) to provide a global evaluation. Then we report the topN-precision and recall@K curves with $k = 64$.

The MAP results achieved by all methods on the ADNI1, ADNI2, and AIBL datasets are reported in Table II. Here, "$M \rightarrow P$" represents the case where queries are sMRI scans and the database contains PET images, and "$P \rightarrow M$" represents the case where queries are PET images and the database has sMRI scans. From the results in Table II, one can observe an interesting finding. That is, the MAP values for "M→P" are usually higher than those for "P→M", especially for ADNI1 and ADNI2. This may be due to the limited number of PET images. From Subsection IV-A, we can see that the numbers of sMRI scans on three datasets are usually larger than that of PET scans. Since deep models usually need a large number of training data to reduce overfitting, the limited PET scans may degrade the generalization ability of the hash functions for PET modality. Therefore, the generated hash codes for MRI queries may be better than those for PET queries, which may induce the performance gap between "M→P" and "P→M" tasks. From the results, one can also obtain that CenterHash usually outperforms other competing methods by large margins. For instance, compared to SCM (the state-of-the-art traditional method), CenterHash can obtain absolute increase of 7.253%/3.105% and 13.74%/4.085% 5.473%/2.915% in terms of the average MAP in two cross-modal retrieval tasks "$M \rightarrow P$" and "$P \rightarrow$

Authorized licensed use limited to: University of North Carolina at Chapel Hill. Downloaded on February 22,2021 at 16:49:06 UTC from IEEE Xplore. Restrictions apply.

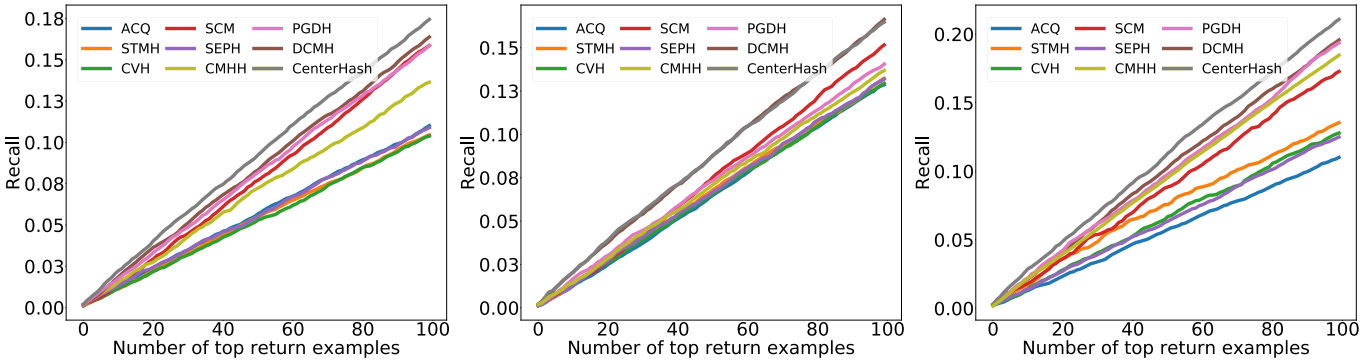
TABLE II
COMPARISON WITH BASELINES IN TERMS OF MAP (WITH BEST RESULTS SHOWN IN BOLDFACE)

Task	Method	ADNI1				ADNI2				AIBL			
		$k=16$	$k=32$	$k=64$	$k=128$	$k=16$	$k=32$	$k=64$	$k=128$	$k=16$	$k=32$	$k=64$	$k=128$
M → P	ACQ	0.4163	0.3713	0.4015	0.4623	0.3883	0.4341	0.4386	0.4096	0.5498	0.5535	0.5563	0.5583
	STMH	0.4246	0.4093	0.4087	0.3864	0.3778	0.4449	0.4511	0.4225	0.5870	0.6087	0.5995	0.5969
	CVH	0.3933	0.3842	0.4163	0.3752	0.4023	0.3896	0.4238	0.4003	0.5694	0.5710	0.5622	0.5654
	SEPH	0.4378	0.4416	0.4063	0.4320	0.4225	0.4610	0.4097	0.4670	0.6231	0.5944	0.5385	0.5660
	SCM	0.5037	0.4713	0.5109	0.5174	0.4368	0.4944	0.4452	0.5009	0.5735	0.5919	0.5892	0.5807
	CMHH	0.4738	0.4323	0.4502	0.4144	0.4520	0.4348	0.4576	0.4223	0.5932	0.5741	0.5794	0.5962
	PGDH	0.5124	0.5181	0.5276	0.5225	0.6275	0.5735	0.5861	0.5543	0.6054	0.6094	0.5928	0.6025
	DCMH	0.5562	0.5529	0.5366	0.5306	0.5819	0.5712	0.5914	0.5573	0.6160	0.6268	0.6261	0.5963
	CenterHash	**0.5855**	**0.5598**	**0.5663**	**0.5818**	**0.6463**	**0.6164**	**0.6363**	**0.6278**	**0.6430**	**0.6361**	**0.6380**	**0.6371**
P → M	ACQ	0.3985	0.4242	0.3725	0.3706	0.3923	0.4023	0.4107	0.4466	0.5544	0.5548	0.5503	0.5503
	STMH	0.4246	0.4093	0.4087	0.3864	0.4281	0.3923	0.3779	0.4586	0.5825	0.5907	0.5731	0.5821
	CVH	0.3913	0.3992	0.4079	0.4138	0.4099	0.3615	0.3768	0.4106	0.6075	0.5880	0.5881	0.5699
	SEPH	0.4585	0.3962	0.4587	0.3688	0.3807	0.3803	0.4223	0.4472	0.5555	0.5651	0.5775	0.5482
	SCM	0.4806	0.4658	0.4599	0.4659	0.4370	0.4853	0.4280	0.4635	0.5798	0.5892	0.5992	0.6106
	CMHH	0.4017	0.4098	0.4072	0.4008	0.4070	0.3961	0.4301	0.4121	0.5888	0.5850	0.5740	0.5964
	PGDH	0.5097	0.4565	0.4422	0.4592	0.4478	0.4585	0.4254	0.4408	0.5848	0.5935	0.5949	0.5996
	DCMH	0.5248	0.4601	0.4518	0.4468	0.4490	0.4521	0.4556	0.4449	0.5923	0.5824	0.5760	0.6118
	CenterHash	**0.5312**	**0.4837**	**0.4989**	**0.4826**	**0.4835**	**0.4970**	**0.5131**	**0.4838**	**0.6251**	**0.6105**	**0.6303**	**0.6295**



(a) ADNI1: M→ P (b) ADNI2: M→ P (c) AIBL: M→ P

Fig. 2. Recall@K curves for ADNI1, ADNI2, and AIBL with sMRI query images and PET database images.



(a) ADNI1: P→ M (b) ADNI2: P→ M (c) AIBL: P→ M

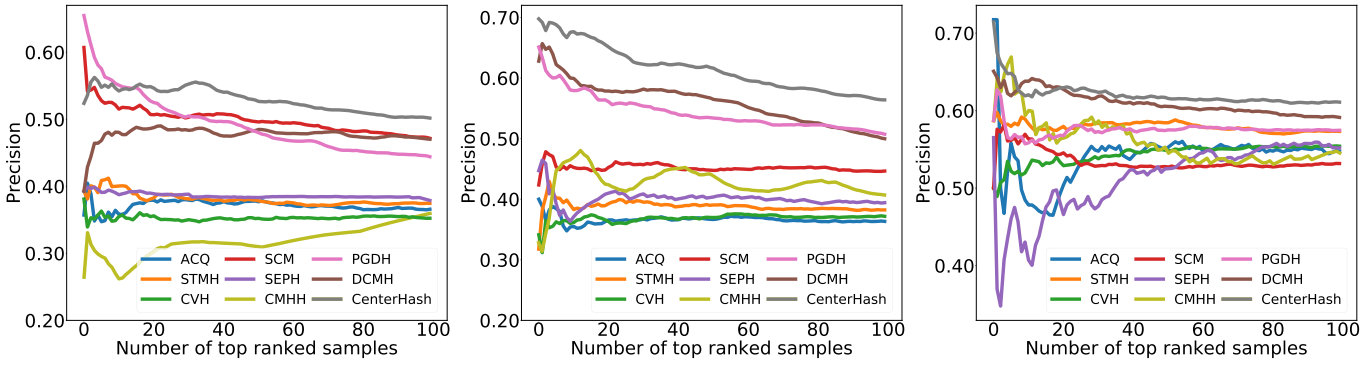
Fig. 3. Recall@K curves for ADNI1, ADNI2, and AIBL with PET query images and sMRI database images.

M" on ADNI1, ADNI2, and AIBL, respectively. Also, CenterHash outperforms DCMH (the state-of-the-art deep learning method), by margins of 2.928%/2.822%, 5.625%/4.395%, and 2.225%/3.323% in terms of the average MAP on three datasets, respectively. These results validate the superiority of CenterHash in cross-modal neuroimage retrieval.

The recall@K curves with $k = 64$ achieved by all methods on the three datasets for both "$M \rightarrow P$" and "$P \rightarrow M$" tasks are shown in Fig. 2 and Fig. 3. From these results, one can obtain that CenterHash can usually achieve better recall performance than other baseline methods on both cross-modal retrieval tasks, which is consistent with the MAP evaluation.

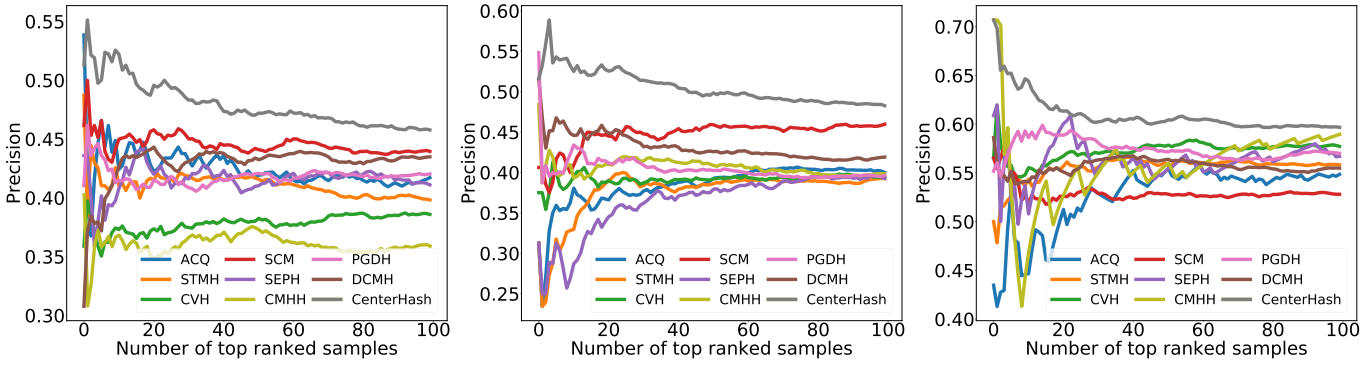
The topN-precision curves with $k = 64$ achieved by different methods on the three datasets for both "$M \rightarrow P$" and "$P \rightarrow M$" tasks are shown in Fig. 4 and Fig. 5. These figures suggest that our CenterHash generally achieves higher precisions, which is consistent with the recall@K and MAP evaluations. In the task of medical image retrieval, users usually focus on the top returned results, and thus, it's essential to provide users with top returned instances that are highly

Authorized licensed use limited to: University of North Carolina at Chapel Hill. Downloaded on February 22,2021 at 16:49:06 UTC from IEEE Xplore. Restrictions apply.



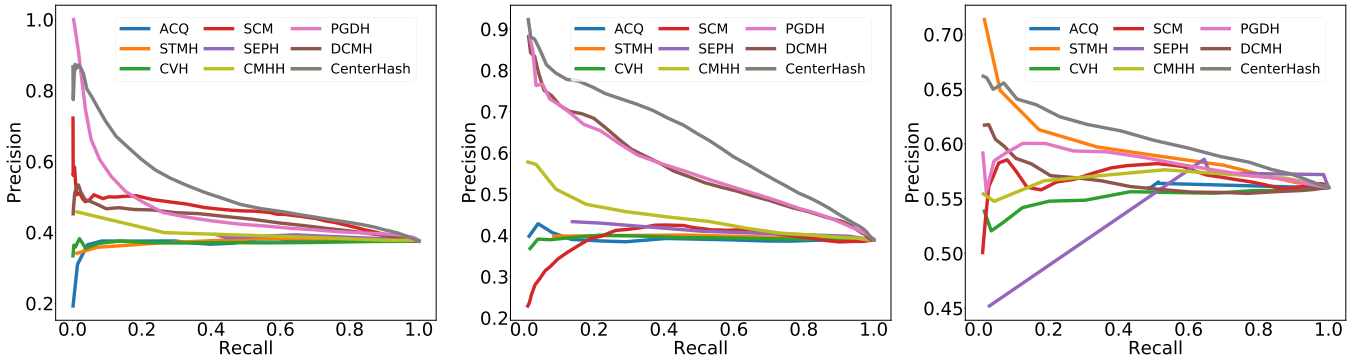
(a) ADNI1: M→ P (b) ADNI2: M→ P (c) AIBL: M→ P

Fig. 4. TopN-precision curves for ADNI1, ADNI2, and AIBL with sMRI query images and PET database images.



(a) ADNI1: P→ M (b) ADNI2: P→ M (c) AIBL: P→ M

Fig. 5. TopN-precision curves for ADNI1, ADNI2, and AIBL with PET query images and sMRI database images.



(a) ADNI1: M→ P (b) ADNI2: M→ P (c) AIBL: M→ P

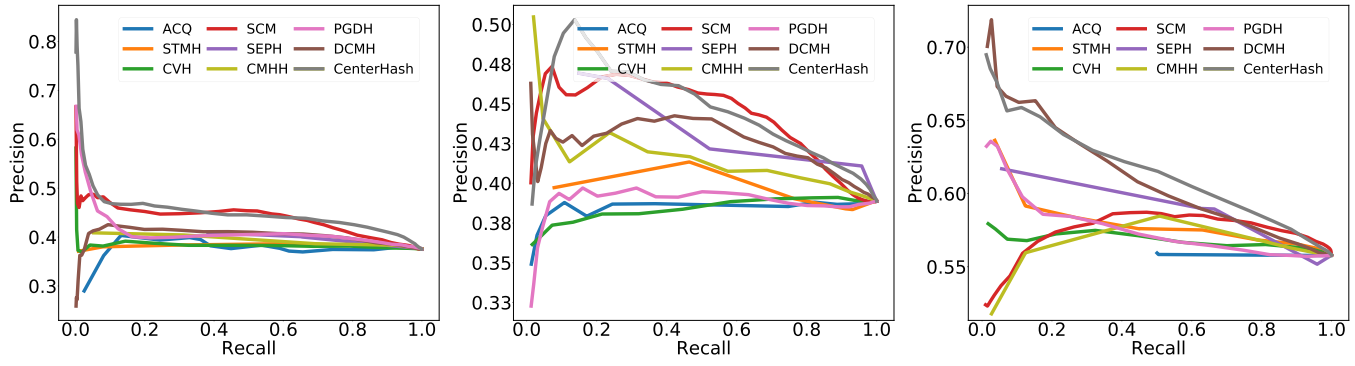
Fig. 6. Precision-recall curves for ADNI1, ADNI2, and AIBL with sMRI query images and PET database images.

relevant to the query. From the results, one can see that CenterHash outperforms the other methods by a large margin when the number of returned instances is small.

2) Results of Hash Lookup: Given a query object, the precision and recall values for the returned items within any Hamming radius can be computed. By investigating these values with every Hamming radius from 0 to k, we can draw the precision-recall curves. Fig. 6 and Fig. 7 report the precision-recall results of different methods on ADNI1, ADNI2, and AIBL for both "$M \rightarrow P$" and "$P \rightarrow M$" tasks with $k = 64$. From these figures, one can observe that CenterHash consistently achieves the best performance. Especially, CenterHash outperforms other methods by a large margin at low recall levels, which is desirable in practical search systems. From Table II and Figs. 2-7, one can see that CenterHash generally obtains superior performance in terms of both Hamming ranking metrics (*i.e.*, MAP, recall@K, and topN-precision) and hash lookup metric (*i.e.*, Precision-recall), which demonstrates that, compared with other baseline methods, the proposed CenterHash can learn more discriminative hash codes and enable more effective neuroimage search on both cross-modal retrieval tasks.

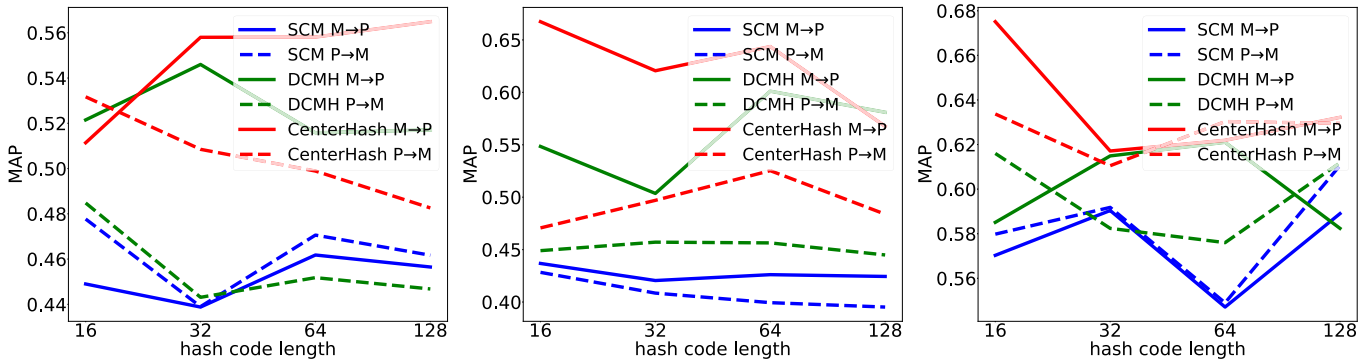
3) Results of Small Hamming Radius: To compare the search performance when the returned examples are in a small Hamming radius, we set Q in Eq. (16) as the number of $k/2$ and report the corresponding MAP results in Fig. 8. From

Authorized licensed use limited to: University of North Carolina at Chapel Hill. Downloaded on February 22,2021 at 16:49:06 UTC from IEEE Xplore. Restrictions apply.



(a) ADNI1: P→ M (b) ADNI2: P→ M (c) AIBL: P→ M

Fig. 7. Precision-recall curves for ADNI1, ADNI2, and AIBL with PET query images and sMRI database images.



(a) ADNI1 (b) ADNI2 (c) AIBL

Fig. 8. MAP results for ADNI1, ADNI2, and AIBL with returned examples in a small Hamming radius.

Fig. 8, one can see that CenterHash can achieve the best MAP results, further demonstrating that CenterHash can successfully learn hash codes with high quality.

D. Cross-Data Generalization

In clinical applications, collecting sufficient medical images for model training may be challenging. Therefore, a retrieval system with strong generalization capabilities across different datasets is highly desirable. Here, to test the generalization ability of CenterHash, we select ADNI1 and ADNI2 datasets and evaluate its search performance in two scenarios: (1) **cross-data training**, where we train the model on ADNI1 (ADNI2) and test it on ADNI2 (ADNI1). The queries and database images are both from ADNI2 (ADNI1); (2) **cross-data query**, where we train the model on ADNI1 (ADNI2) and test its retrieval performance with queries from ADNI2 (ADNI1) and database images from ADNI1 (ADNI2). For both scenarios, two state-of-the-art methods (*i.e.*, SCM and DCMH) are selected for comparison.

1) Cross-Data Training: We first train the models on ADNI1 and test them on ADNI2, with MAP results reported in Fig. 9(a). Similarly, we also train the models on ADNI2 and test them on ADNI1, with results shown in Fig. 9(b). These figures show that our method clearly outperforms SCM and DCMH in both tasks. From Figs. 9(a)-9(b) and Table II, we can observe that CenterHash outperforms or achieves comparable performance with other methods, even though the training and testing images are from different datasets.

2) Cross-Data Query: We first train models on ADNI1. Given query images from ADNI2, we want to retrieval similar subjects in ADNI1. Fig. 10(a) reports the MAP results. Fig. 10(b) shows the results of models trained on ADNI2, where query images are from ADNI1. We can again see that CenterHash consistently achieves the best performance. The results in Figs. 9-10 together demonstrate that our method has strong generalization ability across different datasets.

E. Parameter Sensitivity

We now study the influence of the hyper-parameter α that controls the bandwidth of the adaptive sigmoid function in Eq. (5). Specifically, we first set the hash code length k as 16 and then compute MAP values by varying α between 0.2 and 9.0. The map results achieved by our CenterHash method in the "M→P" and "P→M" tasks on the three datasets are reported in Fig. 11. From this figure, we can see that the MAP values first increase with the increasing of α and then maintain relatively high values in a large range of $1.0 \leq \alpha \leq 5.0$. Thus, in practice, we can empirically select α from $[1.0, 5.0]$. In other experiments in the paper, we fix α as 1.0. The influence of the hash code length can be found in the *Supplementary Materials*.

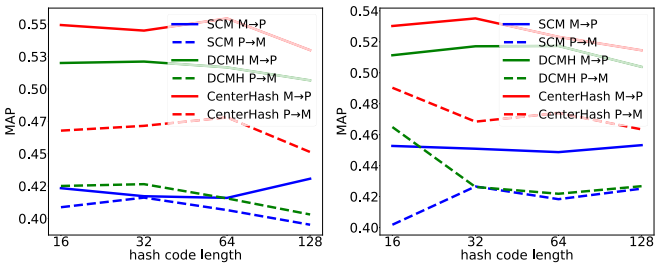
F. Ablation Study

To investigate the role of different components in our framework, we investigate three variants of CenterHash, including (1) **CenterHash-G** that is trained on data from two categories

Authorized licensed use limited to: University of North Carolina at Chapel Hill. Downloaded on February 22,2021 at 16:49:06 UTC from IEEE Xplore. Restrictions apply.

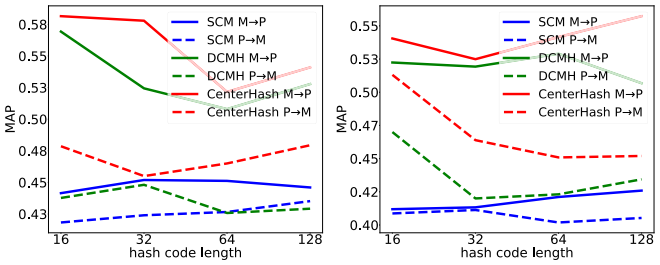
TABLE III
MAP RESULTS OF CENTERHASH AND ITS VARIANTS (WITH BEST RESULTS SHOWN IN BOLDFACE)

Task	Method	ADNI1				ADNI2				AIBL			
		$k=16$	$k=32$	$k=64$	$k=128$	$k=16$	$k=32$	$k=64$	$k=128$	$k=16$	$k=32$	$k=64$	$k=128$
M → P	CenterHash-G	0.4098	0.3848	0.3857	0.3902	0.4371	0.4092	0.4142	0.4046	0.5896	0.5872	0.5794	0.5713
	CenterHash-W	0.5519	0.5433	0.5311	0.5359	0.5933	0.5719	0.6012	0.5760	0.6239	0.6105	0.6043	0.5977
	CenterHash-C	0.5624	0.5292	0.5347	0.5568	0.6191	0.5844	0.5980	0.6039	0.6185	0.6029	0.5994	0.6136
	CenterHash	**0.5855**	**0.5598**	**0.5663**	**0.5818**	**0.6463**	**0.6164**	**0.6363**	**0.6278**	**0.6430**	**0.6361**	**0.6380**	**0.6371**
P → M	CenterHash-G	0.3864	0.3828	0.3621	0.3700	0.4236	0.3972	0.3735	0.3944	0.5898	0.5781	0.5850	0.5771
	CenterHash-W	0.5076	0.4614	0.4537	0.4546	0.4575	0.4775	0.4694	0.4642	0.5849	0.5875	0.5931	0.5863
	CenterHash-C	0.5103	0.4600	0.4517	0.4623	0.4453	0.4608	0.4731	0.4517	0.6038	0.5914	0.5843	0.5889
	CenterHash	**0.5312**	**0.4837**	**0.4989**	**0.4826**	**0.4835**	**0.4970**	**0.5131**	**0.4838**	**0.6251**	**0.6105**	**0.6303**	**0.6295**



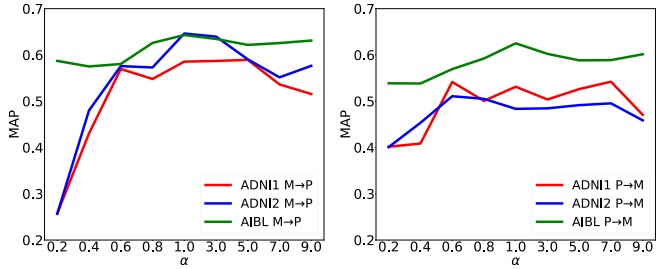
(a) ADNI2@ADNI1 (b) ADNI1@ADNI2

Fig. 9. Results of models using cross-data training strategy. (a) Models trained on ADNI1, and tested on ADNI2 (with query and database images from ADNI2). (b) Models trained on ADNI2, and tested on ADNI1 (with query and database images from ADNI1).



(a) ADNI2→ADNI1 (b) ADNI1→ADNI2

Fig. 10. Results of models using cross-data query strategy. (a) Models trained on ADNI1, and tested on ADNI2 (with query from ADNI2 and database images from ADNI1). (b) Models trained on ADNI2, and tested on ADNI1 (with query from ADNI1 and database images from ADNI2).



(a) M→P (b) P→M

Fig. 11. MAP results on the three datasets with different α.

(i.e., AD and NC) and tested on data from three categories (i.e., AD, NC, and MCI) to test the generalization ability of CenterHash from binary retrieval problems to three-class retrieval problems; (2) **CenterHash-W** that is trained with a standard contrastive likelihood loss [37], [38] (*i.e.*, $w_{ij}=1$ in Eq. (3). This variant is used to demonstrate the role of the weight w_{ij} that is designed to deal with imbalanced training pairs; (3) **CenterHash-C** without using the proposed center prior, which learns hash codes by directly optimizing Eq. (3). This variant is adopted to study the role of the center prior. Note that CenterHash, CenterHash-W and CenterHash-C are trained and tested on data from all three categories (*i.e.*, AD, NC, and MCI). And CenterHash-G is trained on AD and NC subjects and tested on data from those three categories, respectively. The experimental results of CenterHash and its variants are reported in Table III.

From Table III, we can see that CenterHash outperforms **CenterHash-G** in most cases. The underlying reason is that **CenterHash-G**, which is only trained on two categories, cannot capture the data distribution well, resulting in poor generalizability. This also demonstrates the necessity of training models on all three categories. Besides, the performance of **CenterHash-G** on AIBL is generally better than those on ADNI1 and ADNI2. This may be because AIBL contains less MCI images (compared to ADNI1 and ADNI2), so only using AD and NC images for model training has less impact on AIBL. Moreover, the results also show that CenterHash can outperform **CenterHash-W** by substantially large margins of 3.280%/2.978%, 4.610%/2.720% and 2.945%/3.590% in terms of the average MAP for cross-modal retrieval tasks "$M \rightarrow P$" and "$P \rightarrow M$" on ADNI1, ADNI2, and AIBL, respectively. The possible reason is that **CenterHash-W** (using the standard contrastive likelihood loss) neglects the class imbalance problem, and may degrade the retrieval performance. In contrast, CenterHash with the proposed weighted contrastive likelihood can re-weight training pairs according to class importance, providing a flexible solution for the data imbalance problem. Furthermore, we can also obtain that with the proposed center prior, CenterHash outperforms **CenterHash-C** by substantial margins of 2.758%/2.803%, 3.035%/3.663%, and 2.995%/3.175% in terms of the average MAP in the tasks "$M \rightarrow P$" and "$P \rightarrow M$" on ADNI1, ADNI2, and AIBL, respectively. This implies that encouraging hash codes from different modalities to be close to the common class centers (as we do in CenterHash) can potentially enhance the inter-class difference and mitigate the inter-modal discrepancy, thus boosting the search performance.

V. CONCLUSION

This article presents CenterHash for multi-modal neuroimage retrieval. Specifically, the proposed CenterHash formulates the hash code learning problem in a Bayesian learning

Authorized licensed use limited to: University of North Carolina at Chapel Hill. Downloaded on February 22,2021 at 16:49:06 UTC from IEEE Xplore. Restrictions apply.

framework, where a center prior is proposed to explicitly encourage hash codes of the same class from different modalities to be close. Besides, a weighted contrastive likelihood loss function is also developed to address the imbalanced data problem. Comprehensive empirical evidence suggests that our CenterHash yields state-of-the-art cross-modal retrieval performance on three multi-modal neuroimage datasets. As the future work, it is interesting to incorporate the relevance feedback [52] into our framework to further improve the search performance of information retrieval systems.

REFERENCES

[1] N. Tajbakhsh *et al.*, "Convolutional neural networks for medical image analysis: Full training or fine tuning?" *IEEE Trans. Med. Imag.*, vol. 35, no. 5, pp. 1299–1312, May 2016.

[2] C.-H. Cheng and W.-X. Liu, "Identifying degenerative brain disease using rough set classifier based on wavelet packet method," *J. Clin. Med.*, vol. 7, no. 6, p. 124, May 2018.

[3] R. N. J. Graham, R. W. Perriss, and A. F. Scarsbrook, "DICOM demystified: A review of digital file formats and their use in radiological practice," *Clin. Radiol.*, vol. 60, no. 11, pp. 1133–1140, Nov. 2005.

[4] W. E. L. Grimson, R. Kikinis, F. A. Jolesz, and P. Black, "Image-guided surgery," *Sci. Amer.*, vol. 280, no. 6, pp. 54–61, 1999.

[5] S. Gioux, H. S. Choi, and J. V. Frangioni, "Image-guided surgery using invisible near-infrared light: Fundamentals of clinical translation," *Mol. Imag.*, vol. 9, no. 5, p. 7290, 2010.

[6] A. Mohamed, E. Zacharaki, D. Shen, and C. Davatzikos, "Deformable registration of brain tumor images via a statistical model of tumor-induced deformation," *Med. Image Anal.*, vol. 10, no. 5, pp. 752–763, Oct. 2006.

[7] M. Liu, D. Zhang, and D. Shen, "Relationship induced multi-template learning for diagnosis of Alzheimer's disease and mild cognitive impairment," *IEEE Trans. Med. Imag.*, vol. 35, no. 6, pp. 1463–1474, Jun. 2016.

[8] M. Owais, M. Arsalan, J. Choi, and K. R. Park, "Effective diagnosis and treatment through content-based medical image retrieval (CBMIR) by using artificial intelligence," *J. Clin. Med.*, vol. 8, no. 4, p. 462, Apr. 2019.

[9] M. A. P. Purcaru, A. Repanovici, and T. Nedeloiu, "Non-invasive assessment method using thoracic-abdominal profile image acquisition and mathematical modeling with Bezier curves," *J. Clin. Med.*, vol. 8, no. 1, p. 65, Jan. 2019.

[10] U. Avni, H. Greenspan, E. Konen, M. Sharon, and J. Goldberger, "X-ray categorization and retrieval on the organ and pathology level, using patch-based visual words," *IEEE Trans. Med. Imag.*, vol. 30, no. 3, pp. 733–746, Mar. 2011.

[11] M. Liu, J. Zhang, E. Adeli, and D. Shen, "Joint classification and regression via deep multi-task multi-channel learning for Alzheimer's disease diagnosis," *IEEE Trans. Biomed. Eng.*, vol. 66, no. 5, pp. 1195–1206, May 2019.

[12] Y. Fan *et al.*, "Unaffected family members and schizophrenia patients share brain structure patterns: A high-dimensional pattern classification study," *Biol. Psychiatry*, vol. 63, no. 1, pp. 118–124, Jan. 2008.

[13] Y. Fan *et al.*, "Multivariate examination of brain abnormality using both structural and functional MRI," *NeuroImage*, vol. 36, no. 4, pp. 1189–1199, Jul. 2007.

[14] C. P. Slichter, *Principles of Magnetic Resonance*, vol. 1. Berlin, Germany: Springer, 2013.

[15] B. Cheng, M. Liu, D. Shen, Z. Li, and D. Zhang, "Multi-domain transfer learning for early diagnosis of Alzheimer's disease," *Neuroinformatics*, vol. 15, no. 2, pp. 115–132, Apr. 2017.

[16] C. Lian, M. Liu, J. Zhang, and D. Shen, "Hierarchical fully convolutional network for joint atrophy localization and Alzheimer's disease diagnosis using structural MRI," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 42, no. 4, pp. 880–893, Apr. 2020.

[17] D. L. Bailey, M. N. Maisey, D. W. Townsend, and P. E. Valk, *Positron Emission Tomography*. Berlin, Germany: Springer, 2005.

[18] J. Hsieh, *Computed Tomography: Principles, Design, Artifacts, and Recent Advances*, vol. 114. Bellingham, WA, USA: SPIE, 2003.

[19] A. Holt, I. Bichindaritz, R. Schmidt, and P. Perner, "Medical applications in case-based reasoning," *Knowl. Eng. Rev.*, vol. 20, no. 3, pp. 289–292, 2005.

[20] M. Liu, J. Zhang, E. Adeli, and D. Shen, "Landmark-based deep multi-instance learning for brain disease diagnosis," *Med. Image Anal.*, vol. 43, pp. 157–168, Jan. 2018.

[21] Y. Cao *et al.*, "Medical image retrieval: A multimodal approach," *Cancer Informat.*, vol. 13, 2014, p. CIN-S14053.

[22] M. Vikram, A. Anantharaman, and S. Suhas, "An approach for multi-modal medical image retrieval using latent Dirichlet allocation," in *Proc. COMAD*, 2019, pp. 44–51.

[23] Y. Gao, E. Adeli-M, M. Kim, P. Giannakopoulos, S. Haller, and D. Shen, "Medical image retrieval using multi-graph learning for MCI diagnostic assistance," in *Proc. MICCAI*. Munich, Germany: Springer, 2015, pp. 86–93.

[24] J. Song, Y. Yang, Y. Yang, Z. Huang, and H. T. Shen, "Inter-media hashing for large-scale retrieval from heterogeneous data sources," in *Proc. SIGMOD*, 2013, pp. 785–796.

[25] S. Kumar and R. Udupa, "Learning hash functions for cross-view similarity search," in *Proc. IJCAI*, 2011, pp. 1360–1365.

[26] Y. Cao, B. Liu, M. Long, and J. Wang, "Cross-modal Hamming hashing," in *Proc. ECCV*, 2018, pp. 202–218.

[27] Z. Lin, G. Ding, M. Hu, and J. Wang, "Semantics-preserving hashing for cross-view retrieval," in *Proc. CVPR*, Jun. 2015, pp. 3864–3872.

[28] J. Zhang, Y. Peng, and M. Yuan, "Unsupervised generative adversarial cross-modal hashing," in *Proc. AAAI*, 2018, pp. 539–546.

[29] G. Wu *et al.*, "Unsupervised deep hashing via binary latent factor models for large-scale cross-modal retrieval," in *Proc. IJCAI*, Jul. 2018, pp. 2854–2860.

[30] Z. Li, X. Zhang, H. Müller, and S. Zhang, "Large-scale retrieval for medical image analytics: A comprehensive review," *Med. Image Anal.*, vol. 43, pp. 66–84, Jan. 2018.

[31] J. Wang, H. Tao Shen, J. Song, and J. Ji, "Hashing for similarity search: A survey," 2014, *arXiv:1408.2927*. [Online]. Available: http://arxiv.org/abs/1408.2927

[32] C. D. Toth, J. O'Rourke, and J. E. Goodman, *Handbook of Discrete and Computational Geometry*. Boca Raton, FL, USA: CRC Press, 2004.

[33] D. Wang, X. Gao, X. Wang, and L. He, "Semantic topic multimodal hashing for cross-media retrieval," in *Proc. IJCAI*, 2015, pp. 3890–3896.

[34] G. Irie, H. Arai, and Y. Taniguchi, "Alternating co-quantization for cross-modal hashing," in *Proc. ICCV*, Dec. 2015, pp. 1886–1894.

[35] M. M. Bronstein, A. M. Bronstein, F. Michel, and N. Paragios, "Data fusion through cross-modality metric learning using similarity-sensitive hashing," in *Proc. CVPR*, Jun. 2010, pp. 3594–3601.

[36] D. Zhang and W.-J. Li, "Large-scale supervised multimodal hashing with semantic correlation maximization," in *Proc. AAAI*, 2014, p. 7.

[37] Q.-Y. Jiang and W.-J. Li, "Deep cross-modal hashing," in *Proc. CVPR*, Jul. 2017, pp. 3232–3240.

[38] E. Yang, C. Deng, W. Liu, X. Liu, D. Tao, and X. Gao, "Pairwise relationship guided deep hashing for cross-modal retrieval," in *Proc. AAAI Conf. Art. Intell.*, 2017, pp. 1618–1625.

[39] C. Li, C. Deng, N. Li, W. Liu, X. Gao, and D. Tao, "Self-supervised adversarial hashing networks for cross-modal retrieval," in *Proc. CVPR*, Jun. 2018, pp. 4242–4251.

[40] X. Liu, G. Yu, C. Domeniconi, J. Wang, Y. Ren, and M. Guo, "Ranking-based deep cross-modal hashing," in *Proc. AAAI*, 2019, pp. 4400–4407.

[41] V. E. Liong, J. Lu, Y.-P. Tan, and J. Zhou, "Cross-modal deep variational hashing," in *Proc. ICCV*, Oct. 2017, pp. 4077–4085.

[42] W. Liu, J. Wang, R. Ji, Y.-G. Jiang, and S.-F. Chang, "Supervised hashing with kernels," in *Proc. CVPR*, Jun. 2012, pp. 2074–2081.

[43] S. Lloyd, "Least squares quantization in PCM," *IEEE Trans. Inf. Theory*, vol. 28, no. 2, pp. 129–137, Mar. 1982.

[44] C. R. Jack, Jr., *et al.*, "The Alzheimer's disease neuroimaging initiative (ADNI): MRI methods," *J. Magn. Reson. Imag., Off. J. Int. Soc. Magn. Reson. Med.*, vol. 27, no. 4, pp. 685–691, 2008.

[45] K. A. Ellis *et al.*, "The australian imaging, biomarkers and lifestyle (AIBL) study of aging: Methodology and baseline characteristics of 1112 individuals recruited for a longitudinal study of Alzheimer's disease," *Int. Psychogeriatrics*, vol. 21, no. 4, pp. 672–687, Aug. 2009.

[46] J. G. Sled, A. P. Zijdenbos, and A. C. Evans, "A nonparametric method for automatic correction of intensity nonuniformity in MRI data," *IEEE Trans. Med. Imag.*, vol. 17, no. 1, pp. 87–97, Feb. 1998.

[47] Y. Wang, J. Nie, P.-T. Yap, F. Shi, L. Guo, and D. Shen, "Robust deformable-surface-based skull-stripping for large-scale studies," in *Proc. MICCAI*. Springer, 2011, pp. 635–642.

[48] F. Chollet *et al.* (2015). *Keras*. [Online]. Available: https://keras.io

[49] M. Abadi *et al.*, "TensorFlow: A system for large-scale machine learning," in *Proc. OSDI*, vol. 16, 2016, pp. 265–283.

[50] Y. Cao, M. Long, B. Liu, and J. Wang, "Deep cauchy hashing for Hamming space retrieval," in *Proc. CVPR*, Jun. 2018, pp. 1229–1237.

[51] Z. Cao, M. Long, J. Wang, and P. S. Yu, "HashNet: Deep learning to hash by continuation," in *Proc. ICCV*, Oct. 2017, pp. 5608–5617.

[52] Y. Rui, T. S. Huang, and S. Mehrotra, "Content-based image retrieval with relevance feedback in MARS," in *Proc. ICIP*, 1997, pp. 815–818.

Authorized licensed use limited to: University of North Carolina at Chapel Hill. Downloaded on February 22,2021 at 16:49:06 UTC from IEEE Xplore. Restrictions apply.