Cross-Modal Manifold Learning (CM²L) for Medical Image Retrieval

Sailesh Conjeti[1*], Anees Kazi[1*], Nassir Navab[1,4], and Amin Katouzian[3]

[1] Computer Aided Medical Procedures, Technische Universität München, Germany.
[2] IBM Almaden Research Center, San Jose, CA, USA.
[3] Computer Aided Medical Procedures, Johns Hopkins University, USA.

Abstract. This paper presents a new scalable algorithm for cross-modal similarity preserving retrieval in a learnt manifold space. Unlike existing approaches that compromise between preserving global and local geometries, the proposed technique respects both simultaneously during manifold alignment. The global topologies are maintained by recovering underlying mapping functions in the joint manifold space by deploying partially corresponding instances. The inter-, and intra-modality affinity matrices are then computed to reinforce original data skeleton using perturbed minimum spanning tree (pMST), and maximizing the affinity among similar cross-modal instances, respectively. The performance of proposed algorithm is evaluated upon two multimodal image datasets (coronary atherosclerosis histology and brain MRI) for two applications: classification, and regression. Our exhaustive validations and results demonstrate the superiority of our technique over comparative methods and its feasibility for improving computer-assisted diagnosis systems, where disease-specific complementary information shall be aggregated and interpreted across modalities to form the final decision.

1 Introduction

Multi-modal imaging is increasingly performed to obtain disease-specific complementary data throughout diagnosis-treatment-follow-up procedures. A reliable cross-modal image retrieval system is desirable as it carries immense potential in aiding decision-making by enabling access to all information across modalities that share semantic similarity. In contrast to single-modal image retrieval that has been an active research topic [1], the cross modal image retrieval has not yet been fully investigated for medical applications except few works [2], [3], [4] that are also mainly adopted for health care management systems using text+image datasets. In cross-modal retrieval task, the ultimate goal is to bridge the gap between feature spaces by mining their mutual correlations and unveiling similarities within a latent common space. To this end, several methods have been developed but the Canonical Correlation Analysis (CCA) [5] and its variants (ex. [6]) have been widely used for learning such a space by maximizing the correlation between the two feature spaces. Alternatively, learning coupled feature

---

[*] S. Conjeti and A. Kazi contributed equally towards the work.

spaces (LCFS) [7] and Procrustes alignment [8] algorithms have been proposed, where the former focuses on selecting discriminative features while learning the subspace and the latter removes translational, rotational, and scaling components from one space so that the optimal alignment can be achieved. In general, majority of existing methods only preserve local geometries amongst features and ignore global geometries. In other words, they only ensure that similar instances in the original space become neighbors in the latent space but do not prevent dissimilar instances from being neighbors. Authors in [8] addressed this problem by projecting instances into latent space through recovered mapping functions upon partially corresponding instances and aligning the manifolds while preserving the global geometries.

The data across imaging modalities are inherently heterogeneous due to differences in physics of acquisitions and protocols. Therefore, we need to preserve local structures that carry information about population variability and at the same time preserve the global manifold geometries for coping up with disease heterogeneity. This motivated us to develop the cross-modal manifold learning (CM^2L) image retrieval algorithm that respects both geometries in the latent space, which is accounted as the main contribution of this paper. This is achieved by: 1) incorporating pMST [12] into CM^2L such that the original data skeleton is preserved and partially corresponding instances drive the alignment, and 2) introducing novel notion of proximity through inter- and intra-modality affinity matrices for maintaining local similarities within constructed graph neighborhood. To the best of our knowledge, this is the first work on cross-modal medical image retrieval and as proof of concept, we apply CM^2L on two datasets for two different applications. For *classification*, we deployed coronary atherosclerosis histology images stained with Hematoxylin and Eosin (HnE) and Movat Pentachrome (MP). Each stain provides distinctive information about atherosclerotic tissues that are interpreted in search of vulnerable plaques [9], which has high clinical significance. The need for cross-modal retrieval arises when either of two stains are available (*query*) and the histopathologist desires access to the complementary stained image for differential assessment and tissue labeling that is otherwise very tedious and time consuming. Ideally, the retrieved regions of interest should share the same semantic similarity in terms of the disease and arterial morphology as that of the query. For *regression*, we used BraTS dataset, which comprises of 3D scans of 188 High Grade Gliomas (HGG) and 25 Low Grade Gliomas (LGG) acquired with four MRI contrast schemes: T1, TIc, T2, and FLAIR (Refer to [10] for more details). We pose the task of cross-modal retrieval for scenarios where any triad of the above four modalities are available (*query*) and the neurologist seeks the other complementary modality for improved diagnosis and treatment planning. Ideally, the retrieved cross-modal volume/region should share similarity in tumor composition to that of the query triad and similarity in the tumor staging.
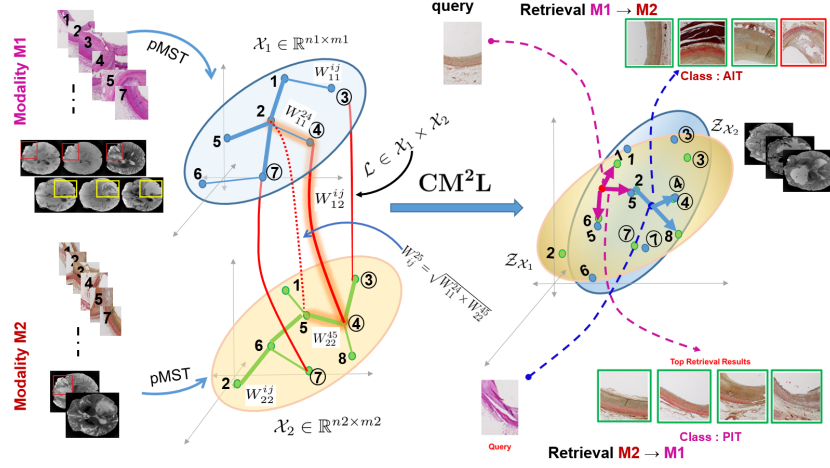
Fig. 1: Retrieval schematic of CM^{2}L: Given modalities $\mathcal{X}_1$, $\mathcal{X}_2$, and limited corresponding instances ($\mathcal{L}$), we model the intra-modal proximity with pMST. We then leverage $\mathcal{L}$ through CM^{2}L and learn to map them to the joint feature space $\mathcal{Z} : \mathcal{Z}_{\mathcal{X}_1} \cup \mathcal{Z}_{\mathcal{X}_2}$.

2 Methodology

Given two sets of multi-modal images $\mathcal{X}_1 = \left\{ \mathbf{x}_1^i \in \mathbb{R}^{m_1} \right\}_{i=1}^{n_1}$ and $\mathcal{X}_2 = \left\{ \mathbf{x}_2^i \in \mathbb{R}^{m_2} \right\}_{i=1}^{n_2}$, as a prerequisite, we collect $n_{\mathcal{L}}$ number of partially corresponding data (tuples) from both modalities constituting $\mathcal{L} \in \mathcal{X}_1 \times \mathcal{X}_2$. We reconstitute $\mathcal{X}_1$ into two disjoint subsets: $\mathcal{X}_1^c$ and $\mathcal{X}_1^{wc}$ (with and without given tuples, respectively) and likewise partition $\mathcal{X}_2$. In presence of limited correspondences, just preserving neighborhood relationships amongst matching instances would result in over-fitting, thus limiting generalization. CM^{2}L overcomes this by using $\mathcal{X}_1^{wc}$ and $\mathcal{X}_2^{wc}$ together with $\mathcal{L}$ such that the whole global geometries of two underlying manifolds are coupled and aligned in the joint feature space.

The CM^{2}L casts the retrieval problem into learning a latent metric q-dimensional space $\mathcal{Z} = (\mathcal{Z}_{\mathcal{X}_1} \cup \mathcal{Z}_{\mathcal{X}_2} | \mathcal{X}_1 \to \mathcal{Z}_{\mathcal{X}_1}; \mathcal{X}_2 \to \mathcal{Z}_{\mathcal{X}_2}) \in R^q$ wherein $\mathcal{X}_1$ and $\mathcal{X}_2$ become comparable ($q \leq \min(m_1, m_2)$).

The overall geometry encompassing both intra- and inter-modal global geometries of the aligning manifolds can be model as a $(n_1 + n_2) \times (n_1 + n_2)$ joint distance matrix $\mathcal{D}$ representing the pairwise dissimilarity between any two instances in $\{\mathcal{X}_1, \mathcal{X}_2\}$. The joint manifold geometry is defined as $\tau(\mathcal{D}) = -HSH/2$, where $S_{ij} = D_{ij}^2$, $H_{ij} = \mathcal{I}^{(n_1+n_2) \times (n_1+n_2)} - (1/(n_1 + n_2))$, and $\mathcal{I}$ is an identity matrix (construction of $\mathcal{D}$ is discussed later) [11]. $\mathcal{Z}$ is estimated through:

$$\left(\mathcal{Z}_{\mathcal{X}_1}^*, \mathcal{Z}_{\mathcal{X}_2}^* \right) = \underset{\mathcal{Z}_{\mathcal{X}_1}, \mathcal{Z}_{\mathcal{X}_2}}{\arg \min} \left\| \tau(\mathcal{D}) - [\mathcal{Z}_{\mathcal{X}_1}, \mathcal{Z}_{\mathcal{X}_2}]^T [\mathcal{Z}_{\mathcal{X}_1}, \mathcal{Z}_{\mathcal{X}_2}] \right\|^2 \qquad (1)$$

Next, the intra- and inter-modal similarities shall be preserved in $\mathcal{Z}$. The former is discovered within modality neighborhoods and modeled as intra-modal

pMSTs. The inter-modality neighborhoods are then inferred using $\mathcal{L}$, which act as 'links' and aid in aligning the intra-modal pMSTs such that similar instances across modalities are mapped close to one another in the learnt manifold space ensuring exclusion of dissimilar points within local neighborhood (see Fig. 1).

`Step 1` $pMST$: The minimum spanning tree (MST) of a data distribution effectively represents the underlying skeleton of a manifold, preserves local structures, and does not introduce *gaps* between small random groupings of data points, which guarantees connectedness of a graph. However, MST is too sparse and sensitive to noise and therefore a fully-connected graph is often constructed, resulting in erroneous connections traveling outside of an underlying manifold when maximizing inter-modality proximity. To resolve this issue, we employed pMST, which is an ensemble of MSTs applied on perturbed versions of the original data distribution $(\mathrm{MST}(\mathcal{X}_p))$, where $\mathcal{X}_p = \{\mathbf{x}_i^p \mid \mathbf{x}_i^p \in \mathcal{N}(\mathbf{x}_i, \sigma_i); \mathbf{x}_i \in \mathcal{X}\}$ and $\sigma_i = r_p \times d(\mathbf{x}_i, \mathbf{x}_i^k)$; with $r_p \in [0,1]$ that is locally adaptive noise model [12]. The edge e_{ij}^p between two points $\mathbf{x}_i^p$ and $\mathbf{x}_j^p$ is 1 if they are connected in $\mathrm{MST}(\mathcal{X}_p)$ and 0 otherwise. The final edge weight between two $\mathbf{x}_i$ and $\mathbf{x}_j$ instances in the pMST neighborhood graph (say $\delta_{\mathcal{X}}$) is computed as $e_{ij} = \frac{1}{t_p} \sum_{p=1}^{t_p} e_{ij}^p$. Next, we define the intra-modal proximity graph, deploying these two constructed pMST graphs, $\delta_{\mathcal{X}_1}$ and $\delta_{\mathcal{X}_2}$ for $\mathcal{X}_1$ and $\mathcal{X}_2$ data points, respectively.

`Step 2` *Intra-modal Affinity*: For defining proper affinity and evaluating similarity/dissimilarity between data points $\mathbf{x}_i$ and $\mathbf{x}_j$ in the manifold, we incorporate locally scaled l_2 norm into intra-modal distance metric $\mathcal{D}_{ij} = \|\mathbf{x}_i - \mathbf{x}_j\|^2 / (2\sigma_{\mathbf{x}_i}\sigma_{\mathbf{x}_j})$, where σ_i and σ_j are local scaling factors measured by $\sigma_i = \|\mathbf{x}_i - \mathbf{x}_K\|^2$ [13] such that $\mathbf{x}_K$ is the Kth neighbor of $\mathbf{x}_i$. This allows for self-tuning of point-to-point distances in local neighborhoods around the points $\mathbf{x}_i$ and $\mathbf{x}_j$. In cross-modal retrieval, the heterogeneous gap between feature spaces warrants that we normalize the intra-modal distance matrices $\mathcal{D}_{11}$ and $\mathcal{D}_{22}$ to make them comparable. Finally, the intra-modal affinity is measured as $W_{11} = \exp(-\mathcal{D}_{11})$. $\delta_{\mathcal{X}_1}$ for $\mathcal{X}_1$ and likewise for $\mathcal{X}_2$.

`Step 3` *Inter-modal Affinity*: Given partial correspondences $\mathcal{L}$, the inter-modal affinity is derived by the inferred intra-modal affinities W_{11}, W_{22}. The corresponding instances across modalities are treated as 'links' to leverage during alignment. For any pair of cross-modal points (say $\mathbf{x}_1^i$ and $\mathbf{x}_2^j$), the cross-modal affinity is computed as the maxima of affinities through all possible 'links' between the modalities, *i.e.* $W_{12}^{ij} = \max_{k \in [1, n_{\mathcal{L}}]} \sqrt{W_{11}^{ik} \times W_{22}^{kj}}$, where W_{11}^{ik} is the intra-modal affinity between $\mathbf{x}_1^i$ and data point $\mathbf{x}_1^k$, and W_{22}^{kj} is the intra-modal affinity between $\mathbf{x}_2^j$ data point $\mathbf{x}_2^k$ where $(\mathbf{x}_1^k, \mathbf{x}_2^k) \in \mathcal{L}$. The final composite distance matrix representing the joint geometry is computed as: $\mathcal{D} = 1 - \begin{pmatrix} W_{11} & W_{12} \\ W_{12}^T & W_{22} \end{pmatrix}$.

Learning Latent Space $\mathcal{Z}$ and Out of Sample Extension: Solution to Eq. 1 posed earlier is the eigen-decomposition of $\tau(\mathcal{D})$ as $\tau(\mathcal{D}) = U^t \mathrm{diag}(\Lambda_1, \cdots, \Lambda_q)U$ where $U \in \mathbb{R}^{(n_1+n_2) \times q}$, and Λ are eigenvalues, therefore, the latent subspace is estimated as $\mathcal{Z} = \mathrm{diag}(\Lambda_1, \cdots, \Lambda_q)^{1/2} U$ [11]. For an unseen data point $\mathbf{x}_t$, we adapt the formulation from [14], which computes locally adaptive tangent spaces

Table 1: Comparative methods and their configurational settings

Methods with abbreviations	Type	Graph	Hyperparameters
Cannonical Correlation Analysis (CCA) [5]	F	×	Cross-modal Correlation > 0.1
Manifold alignment preserving global geometry (MA-F and MA-I) [8]	F and I	FC	Eigen-value threshold $\epsilon > 10E - 05$; k for OSE $= 20$
Learning coupled feature spaces (LCFS) [7]	F	×	Regularization parameters $\lambda_1 = 10E - 01$,$\lambda_2 = 10\text{E-}03$ Number of iterations $= 10$
Procrustus Alignment (PA)	×	×	-
Cross-Modal Manifold Learning (CM^{2}L)	F and I	pMST	Number of perturbations $t_p = 20$ Locally adaptive noise model $r_p = 0.5$; $k = 5$ Eigen-value threshold $\epsilon > 10E - 05$; k for OSE $= 20$

for out of sample extension (OSE). Through OSE, we seek the corresponding projected point $\mathbf{z}_t \in \mathcal{Z}$ and leverage the local neighborhood $N(\mathbf{x}_t)$ defined in the high-dimensional space to search for a locally linear mapping function M such that $\mathbf{z}_t = M\mathbf{x}_t$, where M is decomposable into two piecewise matrices A and V ($M = AV$). We infer V as the eigen-vectors corresponding to the top q non-zero eigenvalues generated through Principal Components Analysis (PCA) on $N(\mathbf{x}_t) \cup \mathbf{x}_t$ and A is the similarity transformation matrix (translation, scaling, and rotation) that is learnt through local Procrustes alignment.

Cross-modal Retrieval in Joint Manifold Space: Through CM^{2}L, we make the projected spaces $\mathcal{Z}_{\mathcal{X}_1}$ and $\mathcal{Z}_{\mathcal{X}_2}$ metric-comparable. Therefore, without loss of generality, the task of cross-modal retrieval for a query (say, $\mathbf{x}_q$ of modality $M1$) will be casted as projecting it appropriately onto the joint space ($\mathbf{z}_q = \text{OSE}(\mathbf{x}_q)$) and fetching the closest projected points from target modality ($\mathcal{Z}_{\mathcal{X}_2}$).

Extension to Feature Level Alignment: So far, the CM^{2}L was elaborated as it searches for and establishes non-linear mapping of original feature spaces and joint embedding space, which we refer to it as "instance-level" version (CM^{2}L-I). It can be seamlessly generalized to the case of linear embedding by replacing $\mathcal{Z}_{\mathcal{X}_1}$ and $\mathcal{Z}_{\mathcal{X}_2}$ in Eq. 1 with $\alpha^t \mathcal{X}_1$ and $\beta^t \mathcal{X}_2$, respectively. The solution is given by the eigenvectors corresponding to the q maximum non-zero eigenvalues of $Z\tau\left(\mathcal{D}\right)V^T\gamma = \lambda VV^T\gamma$ where $V = \begin{pmatrix} \mathcal{X}_1 & 0^{n_1 \times d_2} \\ 0^{n_2 \times d_1} & \mathcal{X}_2 \end{pmatrix}$ where $\gamma = [\alpha, \beta]$ [8]. This linear feature-level variant of CM^{2}L is thereafter refered to as CM^{2}L-F.

3 Experiments and Results

Validation Scheme: The performance of both CM^{2}L-I and CM^{2}L-F algorithms are evaluated against comparative methods as listed in Table 1. We randomly split the data into two disjoint subsets with a 80:20 ratio corresponding to the training and test datasets and repeated the splitting 10 times. For robustness, we quantify the retrieval performance varying the degree of given correspondences for two settings of 20% (sparse) and 80% (dense) correspondences.

Staging of coronary atherosclerosis through cross-modal retrieval: We followed an acquisition protocol in [15] and collected 253 HnE and MP pairs of cross-sections from 16 coronary arteries excised from 6 *post-mortem* human hearts, resulting in 16467 regions of interest (ROIs) with variable sizes (between $640\mu\text{m} \times 640\mu\text{m}$ and $2560\mu\text{m} \times 2560\mu\text{m}$). The stains are performed on consecutive cross-sections ($< 5\mu\text{m}$ apart) and rigidly registered manually. Eleven
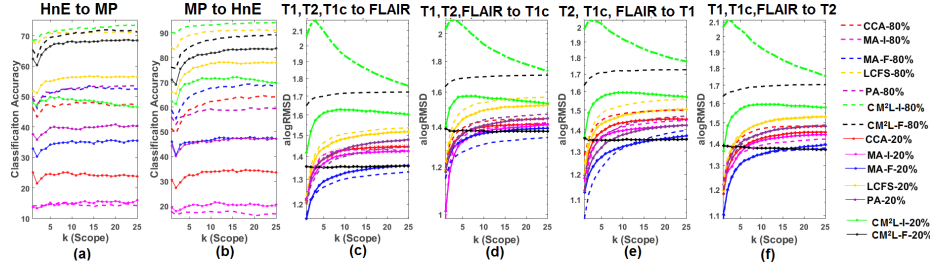
Fig. 2: Performance *vs.* Scope (k retrieved cross-modal neighbors) curves for the proposed and comparative methods for 20% and 80% degrees of correspondence.

Modified AHA [9] labels were used for annotations of underlying tissues in accordance with interpretations from an expert cardiovascular histopathologist. It must be noted that CM^2L does not use labels during training and these annotations are used purely for validations (This extends to BraTS dataset too). The ROIs were then fed into a pre-trained deeply learnt Convolutional Neural Network (CNN) trained for large-scale recognition tasks. We used outputs arising from the second fully connected layer (FC2) of VGG-F [16] and AlexNet [17] deep CNN networks as 4096-dimensional features for HnE and MP images, respectively. Two different networks were chosen to maintain the heterogeneous gap between the raw feature spaces, which would subsequently be bridged through CM^2L and comparative methods. Further, to make the feature spaces discriminative, we reduced dimensionality using supervised locally linear projections, preserving 90% data variance [18].

Results: The retrieval performance are measured using classification accuracy and for a particular query instance (belonging to either HnE or MP) the class is predicted as the maximum a posteriori class evaluated from the top k nearest cross-modal neighbors. Fig. 2 depicts the overall performance, varying k through the accuracy-scope (k) curve for two settings of nearest neighbor retrieval (HnE $\rightarrow$ MP and MP $\rightarrow$ HnE) with 20% and 80% correspondences. Fig. 3(a) depicts the qualitative results of 3 cross-sections and corresponding retrived modality-couterpart images. The normal (N: left column), late fibroatheroma (FA: middle column), and pathological intimal thickening (PIT: right column) plaques have been successfuly retrived on the top 3 ranking results and only two are incorrectly fetched (red boxed) as the fourth neighbors. Such a retrieval tool will significantly improve histopathologist's ability to make reliable and fast decision.

Regression for glioma assessment through cross-modal retrieval: We pose the retrieval task defined over a publicly available multi-protocol MR dataset for Glioma assessment (BraTS) [10]. In total, 2170 variable sized ROIs(> 6 cm $\times$ 6 cm) encompassing tumors were selected across four modalities§. The asso-

---

§ The modalities are pre-aligned to the standard Brainweb space and re-sampled with an isotropic resolution of 1 mm.
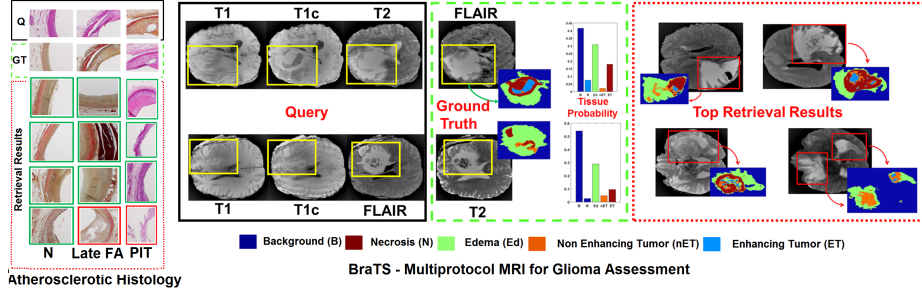
Fig. 3: Qualitative results for (a) atherosclerotic histology and (b) BraTS datasets. Query (Q) image along with 'ground truth' (GT) and top fetched cross-modal images (green box - similar annotation as Query and red box - dissimilar annotation).

ciated ground-truth segmentations are used to construct a five-element tissue probability vector (say, c_p) descriptive of the tumor's composition (in Fig. 3(b)). *Results*: We consider testing scenarios where a triad of modalities are available as a query (say $\mathbf{x}_q$) and the goal is to retrieve the ROIs from the complementary modality that are similar in tissue composition. In this case, the retrieval performance is measured using the average log root mean square deviation (alogRMSD) between the 'ground-truth' tissue probability of the query ($c_p(\mathbf{x}_q)$) to that of the top k nearest cross-modal neighbors over n_t test samples: $\text{alogRMSD} = -\log\left((1/n_t)\sum_{q=1}^{n_t}\sqrt{\sum_{j=1}^{k}\left(c_p(\mathbf{x}_q) - c_p(\mathbf{x}_q^k)\right)^2/k}\right)$. We create four independent test scenarios varying the query triads as: **S1**: T1, T1c, T2 → FLAIR; **S2**: T1, T2, FLAIR → T1c; **S3**: T2, T1c, FLAIR → T1 and **S4**: T1, T1c, FLAIR → T2. Similar to previous experiment, we used AlexNet [17] for generating the query modality features (and naïvely concatenate them), VGG-F [16] for the target modality, and finally removed feature correlations through PCA. Fig. 2 shows the alogRMSD *vs.* scope (k) curves for two comparative settings of 20% and 80% correspondences, evaluating the overall performance. Fig. 3(b) also demonstrates the qualitative retrieval results for two ROIs under **S1** and **S4**, where retrieved cross-modal neighbors (FLAIR(/T2) modality using a T1, T1c, T2(/FLAIR) query triad) exhibit significant tissue compositional similarity with that of the query image. Notably, the expression of necrotic core (with enhancing tumor) engulfed by edema is consistent across retrieved neighbors.

Observations: From Fig. 2, we observe that the two proposed variants of CM^{2}L present a trend of consistently higher performance against comparative methods, substantiating the superiority of preseving joint global and local geometries. The performances of majority of methods are improved as degree of correspondences is increased from 20% to 80%. In case of CM^{2}L, this can be attributed to the better approximation of cross-modal affinity through given corresponding 'links' and hence making cross-modal data comparable in the latent space. Meanwhile, in majority of the cases, CM^{2}L-I shows improved performance over CM^{2}L-F due to the inherent non-linear flexibility of mapping data onto the embedding

space. The LCFS performance is closest to CM^2L as it discovers discriminative common latent space features, which make the embedding compact and effective. Additionally, despite considering global geometries while generating embedding, the MA-I underperformed, because, the Euclidean distance dissimilarity metric is not suitable for representing semantic similarity between instances.

4 Conclusions

We proposed CM^2L for effective cross-modal retrieval in which heterogeneous gap between cross-modal feature spaces is bridged by embedding instances into a metric-comparable latent space. In CM^2L, both local and global geometries are respected simultaneously using limited number of corresponding instances. To the best of our knowledge, this is the first cross-modal medical image retrieval technique and we will extend it to multimedia (text+image) datasets in future.

References

1. Zhang X., *et al.*: Towards Large-Scale Histopathological Image Analysis: Hashing-Based Image Retrieval. IEEE Trans., Med, Imaging, 34(2), 496-506 (2015).
2. Kumar A., *et al.*: A graph-based approach to the retrieval of dual-modality biomedical images using spatial relationships. In EMBC, pp. 390-393 (2008).
3. Cao Y., *et al.*: Medical Image Retrieval: A Multimodal Approach. Cancer Inform, Suppl. 3, pp. 125-36 (2015).
4. Kalpathy-Cramer, J.,*et al.*: Evaluating performance of biomedical image retrieval systems - An overview of the medical image retrieval task at ImageCLEF 2004-2013. Comp. Med. Imag. Grap., 39, 55-61 (2015).
5. Hotelling H.: Relations between two sets of variates. Biometrika, 28:321-77 (1936).
6. Rasiwasia N., *et al.*:A New Approach to Cross-Modal Multimedia Retrieval. ACM Intl Conf Multimedia, pp. 251- 260 (2010).
7. Wang K., *et al.*: Learning coupled feature spaces for cross-modal matching. ICCV, pp. 2088-2095 (2013).
8. Wang C.*et al.*: Manifold Alignment Preserving Global Geometry. IJCAI, p.1-13 (2013).
9. Virmani, *et al.*: Lessons from sudden coronary death a comprehensive morphological classification scheme for atherosclerotic lesions. Arteriosclerosis, thrombosis, and vascular biology, 20(5), pp.1262-1275 (2000)
10. Menze, B. H.,*et al.*: The multimodal brain tumor image segmentation benchmark (BraTS). IEEE TMI, 34(10), 1993-2024 (2015).
11. Tenenbaum, J. B.,*et al.*: A global geometric framework for nonlinear dimensionality reduction. Science, 290(5500), 2319-2323 (2000).
12. Zemel R.S., *et al.* Proximity graphs for clustering and manifold learning. In NIPS, pp. 225-232 (2004).
13. Manor L.Z.*et al.*.: Self-tuning spectral clustering. In NIPS, p. 1601-08 (2004).
14. Strange H., *et al.*: A generalised solution to the out-of-sample extension problem in manifold learning. In AAAI CAI, 1-6 (2011).
15. Katouzian, A.,*et al.*: Methods in Atherosclerotic Plaque Characterization Using Intravascular Ultrasound Images and Backscattered Signals. In Atherosclerosis Disease Management, pp. 121-152 (2011).

16. Chatfield, K.,*et al.*: Return of the devil in the details: Delving deep into convolutional nets. arXiv preprint arXiv:1405.3531 (2014).
17. Krizhevsky, A.,*et al.*: Imagenet classification with deep convolutional neural networks. In NIPS, p.1097-1105 (2012).
18. He, X., & Niyogi, P.: Locality Preserving Projections. In NIPS, pp. 153-160 (2004).