Deep Multiple Instance Hashing for Scalable Medical Image Retrieval

Sailesh Conjeti[1(✉)], Magdalini Paschali[1], Amin Katouzian[2],
and Nassir Navab[1,3]

[1] Computer Aided Medical Procedures, Technische Universität München,
Munich, Germany
sailesh.conjeti@tum.de
[2] IBM Almaden Research Center, Almaden, USA
[3] Computer Aided Medical Procedures, Johns Hopkins University, Baltimore, USA

Abstract. In this paper, for the first time, we introduce a *multiple instance* (MI) deep hashing technique for learning discriminative hash codes with weak bag-level supervision suited for large-scale retrieval. We learn such hash codes by aggregating deeply learnt hierarchical representations across bag members through an MI pool layer. For better trainability and retrieval quality, we propose a two-pronged approach that includes robust optimization and training with an auxiliary single instance hashing arm which is down-regulated gradually. We pose retrieval for tumor assessment as an MI problem because tumors often coexist with benign masses and could exhibit complementary signatures when scanned from different anatomical views. Experimental validations demonstrate improved retrieval performance over the state-of-the-art methods.

1 Introduction

In breast examinations, such as mammography, detected actionable tumors are further examined through invasive histology. Objective interpretation of these modalities is fraught with high inter-observer variability and limited reproducibility [1]. In this context, a reference based assessment, such as presenting prior cases with similar disease manifestations (termed Content Based Image Retrieval (CBIR)) could be used to circumvent discrepancies in cancer grading. With growing sizes of clinical databases, such a CBIR system ought to be both scalable and accurate. Towards this, hashing approaches for CBIR are being actively investigated for representing images as compact binary codes that can be used for fast and accurate retrieval [2–4].

Malignant carcinomas are often co-located with benign manifestations and suspect normal tissues [5]. In such cases, describing the whole image with a single label is inadequate for objective machine learning and alternatively requires expert annotations delineating the exact location of the region of interest.

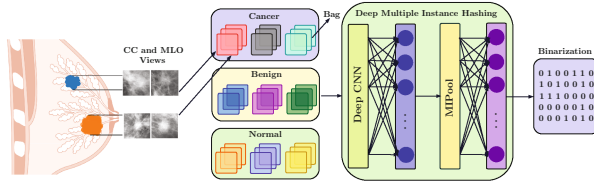S. Conjeti and M. Paschali are contributed equally.

Fig. 1. Overview of DMIH for end-to-end generation of bag-level hash codes. Breast anatomy image is attributed to Cancer Research UK/Wikimedia Commons.

This argument extends to screening modalities like mammograms, where multiple anatomical views are acquired. In such scenarios, the status of the tumor is best represented to a CBIR system by constituting a bag of all associated images, thus veritably becoming multiple instance (MI) in nature. With this as our premise we present, for the first time, a novel deep learning based MI hashing method, termed as Deep Multiple Instance Hashing (DMIH).

Seminal works on shallow learning-based hashing include Iterative Quantization (ITQ) [6], Kernel Sensitive Hashing (KSH) [2] *etc.* that propose a two-stage framework involving extraction of hand-crafted features followed by binarization. Yang *et al.* extend these methods to MI learning scenarios with two variants: Instance Level MI Hashing (IMIH) and Bag Level MI Hashing (BMIH) [7]. However, these approaches are not end-to-end and are susceptible to semantic gap between features and associated concepts. Alternatively, deep hashing methods such as simultaneous feature learning and hashing (SFLH) [8], deep hashing networks (DHN) [9] and deep residual hashing (DRH) [3] propose the learning of representations and hash codes in an end-to-end fashion, in effect bridging this semantic gap. It must be noted that all the above deep hashing works targeted single instance (SI) hashing scenarios and an extension to MI hashing was not investigated.

Earlier works on MI deep learning in computer vision include work by Wu *et al.* [10], where the concept of an MI pooling (MIPool) layer is introduced to aggregate representations for multi-label classification. Yan *et al.* leveraged MI deep learning for efficient body part recognition [11]. Unlike MI classification that potentially substitutes the decision of the clinician, retrieval aims at presenting them with richer contextual information to facilitate decision-making.

DMIH effectively bridges the two concepts for CBIR systems by combining the representation learning strength of deep MI learning with the potential for scalability arising from hashing. Within CBIR for breast cancer, notable prior art includes work on mammogram image retrieval by Jiang *et al.* [12] and large-scale histology retrieval by Zhang *et al.* [4]. Both these works pose CBIR as an SI retrieval problem. Contrasting with [4,12], within DMIH we create a bag of images to represent a particular pathological case and generate a bag-level hash code, as shown in Fig. 1. Our contributions in this paper include: **(1)** introduction of a robust supervised retrieval loss for learning in presence of weak labels and potential outliers; **(2)** training with an auxiliary SI arm with gradual loss

trade-off for improved trainability; and **(3)** incorporation of the MIPool layer to aggregate representations across variable number of instances within a bag, generating bag-level discriminative hash codes.

2   Methodology

Lets consider database $\mathcal{B} = \{B_1, \ldots, B_{N_B}\}$ with N_B bags. Each bag, B_i, with varying number (n_i) of instances (I_i) is denoted as $B_i = \{I_1, \ldots, I_{n_i}\}$. We aim at learning $\mathcal{H}$ that maps each bag to a K-d Hamming space $\mathcal{H} : \mathcal{B} \rightarrow \{-1, 1\}^K$, such that *bags with similar instances and labels are mapped to similar codes*. For supervised learning of $\mathcal{H}$, we define a bag-level pairwise similarity matrix $\mathcal{S}^{\mathrm{MI}} = \{s_{ij}\}_{ij=1}^{N_B}$, such that $s_{ij} = 1$ if the bags are similar and zero otherwise. In applications, such as this one, where retrieval ground truth is unavailable we can use classification labels as a surrogate for generating $\mathcal{S}^{\mathrm{MI}}$.

Architecture: As shown in Fig. 2, the proposed DMIH framework consists of a deep CNN terminating in a fully connected layer (FCL). Its outputs $\{z_{ij}\}_{j=1}^{n_i}$ are fed into the MIPool layer to generate the aggregated representation $\hat{z}_i$ that is pooled ($\max_{\forall j} \{z_{ij}\}_{j=1}^{n_i}$, mean $(\cdot)$, *etc.*) across instances within the bag. $\hat{z}_i$ is an embedding in the space of the bags and is the input of a fully connected MI hashing layer. The output of this layer is squashed to $[-1, 1]$ by passing it through a $\tanh\{\cdot\}$ function to generate h_i^{MI}, which is quantized to produce bag-level hash codes as $\mathbf{b}_i^{\mathrm{MI}} = \mathrm{sgn}\,(\mathbf{h}_i^{\mathrm{MI}})$. The deep CNN mentioned earlier could be a pretrained network, such as VGGF [13], GoogleNet [14], ResNet50 (R50) [15] or an application specific network.

Fig. 2. DMIH architecture.

During training of DMIH, we introduce an auxiliary SI hashing (aux-SI) arm, as shown in Fig. 2. It taps off at the FCL layer and feeds directly into a fully connected SI hashing layer with $\tanh\{\cdot\}$ activation to generate instance level non-quantized hash codes, denoted as $\{h_{ij}^{\mathrm{SI}}\}_{j=1}^{n_i}$. While training DMIH using backpropagation, the MIPool layer significantly sparsifies the gradients (analogous to using very high dropout while training CNNs), thus limiting the trainability of the preceding layers. The SI hashing arm helps to mitigate this by producing auxiliary instance level gradients.

Model Learning and Robust Optimization: To learn similarity preserving hash codes, we propose a robust version of supervised retrieval loss based on neighborhood component analysis (NCA) employed by [16]. The motivation to introduce robustness within the loss function is two-fold: (1) robustness induces immunity to potentially noisy labels due to high inter-observer variability and limited reproducibility for the applications at hand [1]; (2) it can
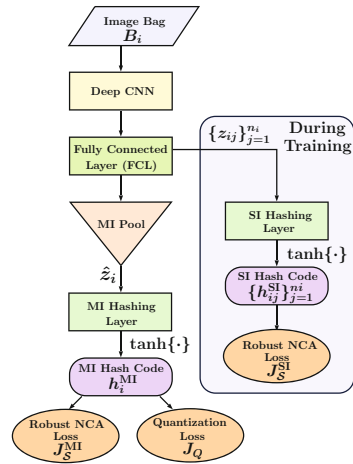
effectively counter ambiguous label assignment while training with the aux-SI hashing arm. Given $\mathcal{S}^{\mathrm{MI}}$, the robust supervised retrieval loss $J_{\mathcal{S}}^{\mathrm{MI}}$ is defined as: $J_{\mathcal{S}}^{\mathrm{MI}} = 1 - \frac{1}{N_B^2} \sum_{i,j=1}^{N_B} s_{ij} p_{ij}$ where p_{ij} is the probability that two bags (indexed as i and j) are neighbors. Given hash codes $\mathbf{h_i} = \{h_i^k\}_{k=1}^{K}$ and $\mathbf{h_j}$, we define a bit-wise residual operation r_{ij} as $r_{ij}^k = (h_i^k - h_j^k)$. We estimate p_{ij} as:

$$p_{ij} = \frac{e^{-\mathcal{L}_{\mathrm{Huber}}(\mathbf{h_i}, \mathbf{h_j})}}{\sum_{i \neq l}^{N_B} e^{-\mathcal{L}_{\mathrm{Huber}}(\mathbf{h_i}, \mathbf{h_l})}}, \text{ where } \mathcal{L}_{\mathrm{Huber}}(\mathbf{h_i}, \mathbf{h_j}) = \sum_{\forall k} \rho_k(r_{ij}^k). \tag{1}$$

$\mathcal{L}_{\mathrm{Huber}}(\mathbf{h_i}, \mathbf{h_j})$ is the Huber norm between hash codes for bags i and j, while the robustness operation ρ_k is defined as:

$$\rho_k(r_{ij}^k) = \begin{cases} \frac{1}{2}(r_{ij}^k)^2, & \text{if } |r_{ij}^k| \leqslant c_k \\ c_k|r_{ij}^k| - \frac{1}{2}c_k^2, & \text{if } |r_{ij}^k| > c_k \end{cases} \tag{2}$$

In Eq. (2), the tuning factor c_k is estimated inherently from the data and is set to $c_k = 1.345 \times \sigma_k$. The factor of 1.345 is chosen to provide approximately 95% asymptotic efficiency and σ_k is a robust measure of bit-wise variance of r_{ij}^k. Specifically, σ_k is estimated as 1.485 times the median absolute deviation of r_{ij}^k as empirically suggested in [17]. This robust formulation provides immunity to outliers during training by clipping their gradients. For training with the aux-SI hashing arm, we employ a similar robust retrieval loss $J_{\mathcal{S}}^{\mathrm{SI}}$ defined over single instances with bag-labels assigned to member instances.

To minimize loss of retrieval quality due to quantization, we use a differentiable quantization loss $J_Q = \sum_{i=1}^{M}(\log \cosh(|\mathbf{h_i}| - \mathbf{1}))$ proposed in [9]. This loss also counters the effect of using continuous relaxation in definition of p_{ij} over using Hamming distance. As a standard practice in deep learning, we also add an additional weight decay regularization term R_W, which is the Frobenius norm of the weights and biases, to regularize the cost function and avoid over-fitting. The following composite loss is used to train DMIH:

$$J = \lambda_{\mathrm{MI}}^t J_{\mathcal{S}}^{\mathrm{MI}} + \lambda_{\mathrm{SI}}^t J_{\mathcal{S}}^{\mathrm{SI}} + \lambda_q J_Q + \lambda_w R_W \tag{3}$$

where λ_{MI}^t, λ_{SI}^t, λ_q and λ_w are hyper-parameters that control the contribution of each of the loss terms. Specifically, λ_{MI}^t and λ_{SI}^t control the trade-off between the MI and SI hashing losses. The SI arm plays a significant role only in the early stages of training and can be traded off eventually to avoid sub-optimal MI hashing. For this we introduce a weight trade-off formulation that gradually down-regulates λ_{SI}^t, while simultaneously up-regulating λ_{MI}^t. Here, we use $\lambda_{\mathrm{SI}}^t = 1 - 0.5\left(1 - t/t_{\max}\right)^2$ and $\lambda_{\mathrm{MI}}^t = 1 - \lambda_{\mathrm{SI}}^t$, where t is the current epoch and $t_{\max}$ is the maximum number of epochs (see Fig. 3). We train DMIH with mini-batch stochastic gradient descent (SGD) with momentum. Due to potential outliers
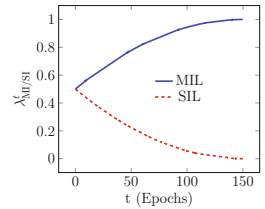

Fig. 3. Weight trade-off.

that can occur at the beginning of training, we scale c_k up by a factor of 7 for $t = 1$ to allow a stable state to be reached.

3   Experiments

Databases: Clinical applicability of DMIH has been validated on two large scale datasets, namely, Digital Database for Screening Mammography (DDSM) [12,18] and a retrospectively acquired histology dataset from the Indiana University Health Pathology Lab (*IUPHL*) [4,19]. The *DDSM* dataset comprises of 11,617 expert selected regions of interest (ROI) curated from 1861 patients. Multiple ROIs associated with a single breast from two anatomical views constitute a bag (size: 1–12; median: 2), which has been annotated as normal, benign or malignant by expert radiologists. A bag labeled *malignant* could potentially contain multiple suspect normal and benign masses, which have not been individually identified. The *IUPHL* dataset is a collection of 653 ROIs from histology slides from 40 patients (20 with precancerous ductal hyperplasia (UDH) and rest with ductal carcinoma *in situ* (DCIS)) with ROI level annotations done by expert histopathologists. Due to high variability in sizes of these ROIs (upto $9\,\mathrm{K} \times 8\,\mathrm{K}$ pixels), we extract multiple patches and populate a ROI-level bag (size: 1–15; median: 8). From both the datasets, we use patient-level non-overlapping splits to constitute the training (80%) and testing (20%) sets.

Model Settings and Validations: To validate proposed contributions, namely robustness within NCA loss and trade-off from the aux-SI arm, we perform ablative testing with combinations of their baseline variants by fine-tuning multiple network architectures. Additionally, we compare DMIH against four state-of-the art methods: ITQ [6], KSH [2], SFLH [8] and DHN [9]. For a fair comparison, we use R50 for both SFLH and DHN, since as discussed later it performs the best. Since SFLH and DHN were originally proposed for SI hashing, we introduce additional MI variants by hashing through the MIPool layer. For ITQ and KSH, we further create two comparative settings: **(1)** Using IMIH [7] that learns instance-level hash codes followed by bag-level distance computation and **(2)** Utilizing BMIH [7] using bag-level kernalized representations followed by binarization.

For IMIH and SI variants of SFLH, DHN and DMIH, given two bags B_p and B_q with SI hash codes, say $\mathcal{H}(B_q) = \{h_{q1}, \ldots, h_{qM}\}$ and $\mathcal{H}(B_p) = \{h_{p1}, \ldots, h_{pN}\}$, the bag-level distance is computed as:

$$d(B_p, B_q) = \frac{1}{M} \sum_{i=1}^{M} (\min_{\forall j} \; \mathrm{Hamming}(h_{pi}, h_{qj})). \tag{4}$$

All images were resized to 224×224 and training data were augmented with random rigid transformations to create equally balanced classes. λ_{MI}^t and λ_{SI}^t were set assuming $t_{\max}$ as 150 epoch; λ_q and λ_w were set at 0.05 and 0.001 respectively. The momentum term within SGD was set to 0.9 and batch size to 128 for *DDSM* and 32 for *IUPHL*. For efficient learning, we use an exponentially decaying learning rate initialized at 0.01. The DMIH framework was
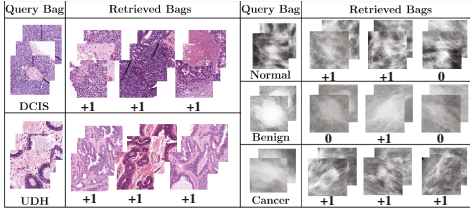
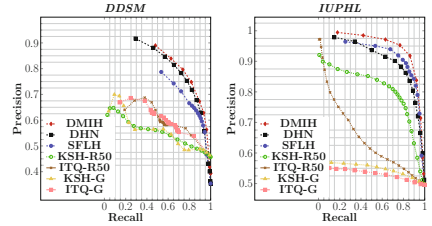**Fig. 4.** Retrieval results for DMIH at code size 16 bits.

Fig. 5. PR curves for DDSM and IUPHL datasets at code size of 32.

Table 1. Performance of ablative testing at code size of 16 bits. We report the nearest neighbor classification accuracy (nnCA) estimated over unseen test data. Letters A-E are introduced for easier comparisons, discussed in Sect. 4.

Method		Variants		DDSM			IUPHL		
		R	T	VGGF	R50	GN	VGGF	R50	GN
Ablative Testing	A	○	○	68.65	72.76	71.70	83.85	85.42	82.29
	B	○	●	75.38	77.34	72.92	85.94	90.10	88.02
	C	●	○	70.65	76.63	70.02	83.33	85.94	86.46
	D	○	■	66.65	69.67	68.26	83.33	88.54	84.90
	E	●	■	67.05	76.59	72.84	84.38	89.58	85.42
DMIH-mean		●	●	78.67	82.31	76.83	87.50	89.58	**89.06**
DMIH-max		●	●	**81.21**	**85.68**	**78.67**	**91.67**	**95.83**	88.02
DMIH($\lambda_q = 0$)		●	●	75.34	79.88	73.06	87.50	89.58	88.51
DMIH NB		●	●	83.25	88.02	79.06	94.79	96.35	92.71
Legend	**R(Robustness)**			○ = L_2, ● = L_{Huber}					
	T(Trade-off)			○ = Equal weights, ● = Decaying SIL weights, ■ = No SIL branch					
	Networks			R50: ResNet50, GN: GoogleNet					

implemented in MatConvNet [20]. We use standard retrieval quality metrics: nearest neighbor classification accuracy (nnCA) and precision-recall (PR) curves to perform the aforementioned comparisons. The results (nnCA) from ablative testing and comparative methods are tabulated in Tables 1 and 2 respectively. Within Table 2, methods were evaluated at two different code sizes (16 bits and 32 bits). We also present the PR curves of select bag-level methods (32 bits) in Fig. 5.

4   Results and Discussion

Effect of aux-SI Loss: To justify using the aux-SI loss, we introduce a variant of DMIH without it (E in Table 1), which leads to a significant decline of 3% to 14% in contrast to DMIH. This could be potentially attributed to the prevention of the gradient sparsification caused by the MIPool layer. From Table 1, we observe a 3%–10% increase in performance, comparing cases with gradual decaying trade-off (B) against baseline setting ($\lambda_{\text{MI}}^t = \lambda_{\text{SI}}^t = 0.5$, A, C).

Effect of Robustness: For robust-NCA, we compared against the original NCA formulation proposed in [16] (A, B, D in Table 1). Robustness helps handle potentially noisy MI labels, inconsistencies within a bag and the ambiguity in

assigning SI labels. Comparing the effect of robustness for baselines sans the SI hashing arm (D *vs.* E) we observe marginally positive improvement across the architectures and datasets, with a substantial 7% in ResNet50 for *DDSM*. Robustness contributes more with the addition of the aux-SI hash arm (proposed *vs.* E) with improved performance in the range of 4%–5% across all settings. This observation further validates our prior argument.

Effect of Quantization: To assess the effect of quantization, we define two baselines: (1) setting $\lambda_q = 0$ and (2) using non-quantized hash codes for retrieval (DMIH - NB). The latter potentially acts as an upper bound for performance evaluation. From Table 1, we observe a consistent increase in performance by margins of 3%–5% if DMIH is learnt with an explicit quantization loss to limit the associated error. It must also be noted that comparing with DMIH - NB, there is only a marginal fall in performance (2%–4%), which is desired.

As a whole, the two-pronged proposed approach, including robustness and trade-off, along with quantization loss delivers the highest performance, proving that DMIH is able to learn effectively, despite the ambiguity induced by the SI hashing arm. Figure 4 demonstrates the retrieval performance of DMIH on the target databases. For *IUPHL*, the retrieved images are semantically similar to the query as consistent anatomical signatures are evident in the retrieved neighbors. For *DDSM*, in the cancer and normal cases the retrieved neighbors are consistent, however it is hard to distinguish between benign and malignant. The retrieval time for a single query for DMIH was observed at 31.62 ms (for *IUPHL*) and 17.48 ms (for *DDSM*) showing potential for fast and scalable search.

Table 2. Results of comparison with state-of-the art hashing methods.

Method		A/F	L	DDSM		IUPHL	
				16-bit	32-bit	16-bit	32-bit
Shallow	ITQ [6]	R50	○	66.35	67.71	78.58	80.28
	KSH [2]	R50	●	64.56	71.98	89.58	79.69
		G	○	65.22	66.55	51.79	51.42
		G	●	59.73	61.03	57.29	58.85
		R50	○	61.88	64.81	87.74	86.51
		R50	●	59.81	72.17	70.83	80.21
		G	○	60.50	61.91	57.36	57.83
		G	●	55.34	55.67	60.94	58.85
Deep	SFLH [8]	R50	○	73.54	77.46	83.33	85.94
		R50M	■	71.98	75.93	85.42	88.54
	DHN [9]	R50	○	65.64	74.79	82.29	86.46
		R50M	■	72.88	80.43	88.02	90.62
	DMIH-SIL	R50	○	76.02	78.37	87.92	88.58
	DMIH	R50M	■	**85.68**	**89.47**	**95.83**	**93.23**
Legend	**A/F:**	**A**: Architecture, **F**: Features R50: ResNet50, R50M: ResNet50+MIPool, G: GIST					
	L:	○ = IMIH, ● = BMIH, ■ = End-to-end					

Comparative Methods

In the contrastive experiments against ITQ and KSH, hand-crafted GIST [21] features under-performed significantly, while the improvement with the R50 features ranged from 5%–30%. However, DMIH still performed 10%–25% better.

Comparing the SI with the MI variations of DHN, SFLH and DMIH, it is observed that the performance improved in the range of 3%–11%, suggesting that end-to-end learning of MI hash codes is preferred over two-stage hashing *i.e.* hashing at SI level and comparing at bag level with Eq. (4). However, DMIH fares comparably better than both the SI and MI versions of SFLH and DHN, owing to the robustness of the proposed

retrieval loss function. As also seen from the associated PR curves in Fig. 5, the performance gap between shallow and deep hashing methods remains significant despite using R50 features. Comparative results strongly support our premise that end-to-end learning of MI hash codes is preferred over conventional two-stage approaches.

5   Conclusion

In this paper, for the first time, we propose an end-to-end deep robust hashing framework, termed DMIH, for retrieval under a multiple instance setting. We incorporate the MIPool layer to aggregate representations across instances to generate a bag-level discriminative hash code. We introduce the notion of robustness into our supervised retrieval loss and improve the trainability of DMIH by utilizing an aux-SI hashing arm regulated by a trade-off. Extensive validations and ablative testing on two public breast cancer datasets demonstrate the superiority of DMIH and its potential for future extension to other MI applications.

References

1. Duijm, L.E.M., Louwman, M.W.J., Groenewoud, J.H., van de Poll-Franse, L.V., Fracheboud, J., Coebergh, J.W.: Inter-observer variability in mammography screening and effect of type and number of readers on screening outcome. BJC **24**, 901–907 (2009)
2. Liu, W., Wang, J., Ji, R., Jiang, Y.G., Chang, S.F.: Supervised hashing with kernels. In: CVPR 2012, pp. 2074–2081. IEEE (2012)
3. Conjeti, S., Guha Roy, A., Katouzian, A., Navab, N.: Hashing with residual networks for image retrieval. In: 20th International Conference on Medical Image Computing and Computer Assisted Intervention, Canada (2017)
4. Zhang, X., Liu, W., Dundar, M., Badve, S., Zhang, S.: Towards large-scale histopathological image analysis: hashing-based image retrieval. In: TMI 2015. IEEE (2015)
5. Veta, M., Pluim, J.P., van Diest, P.J., Viergever, M.A.: Breast cancer histopathology image analysis: a review. Trans. Biomed. Eng. **61**, 1400–1411 (2014). IEEE
6. Gong, Y., Lazebnik, S.: Iterative quantization: a procrustean approach to learning binary codes. In: CVPR 2011, pp. 817–824. IEEE (2011)
7. Yang, Y., Xu, X.-S., Wang, X., Guo, S., Cui, L.: Hashing multi-instance data from bag and instance level. In: Cheng, R., Cui, B., Zhang, Z., Cai, R., Xu, J. (eds.) APWeb 2015. LNCS, vol. 9313, pp. 437–448. Springer, Cham (2015). doi:10.1007/978-3-319-25255-1_36
8. Lai, H., Pan, Y., Liu, Y., Yan, S.: Simultaneous feature learning and hash coding with deep neural networks. In: CVPR 2015, pp. 3270–3278 (2015)

9. Zhu, H., Long, M., Wang, J., Cao, Y.: Deep hashing network for efficient similarity retrieval. In: AAAI 2016 (2016)
10. Wu, J., Yu, Y., Huang, C., Yu, K.: Multiple instance learning for image classification and auto-annotation. In: CVPR 2015, pp. 3460–3469 (2015)
11. Zhennan, Y., Yiqiang, Z., Zhigang, P., Shu, L., Shinagawa, Y., Shaoting, Z., Metaxas, D.N., Xiang, S.Z.: Multi-instance deep learning: discover discriminative local anatomies for bodypart recognition. Trans. Med. Imaging **35**, 1332–1343 (2016). IEEE
12. Jiang, M., Zhang, S., Li, H., Metaxas, D.N.: Computer-aided diagnosis of mammographic masses using scalable image retrieval. TBME **62**, 783–792 (2015)
13. Chatfeld, K., Simonyan, K., Vedaldi, A., Zisserman, A.: Return of the devil in the details: delving deep into convolutional nets (2014). arXiv:1405.3531
14. Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S.E., Anguelov, D., Erhan, D., Vanhoucke, V., Rabinovich, A.: Going deeper with convolutions. In: CVPR 2015, pp. 1–9 (2015)
15. He, K., Zhang, X., Ren, S., Sun, J.: Deep residual learning for image recognition. In: CVPR 2016, pp. 770–778. IEEE Computer Society (2016)
16. Torralba, A., Fergus, R., Weiss, Y.: Small codes and large image databases for recognition. In: CVPR 2008, pp. 1–8. IEEE (2008)
17. Huber, P.J.: Robust statistics. In: Lovic, M. (ed.) International Encyclopedia of Statistical Science, pp. 1248–1251. Springer, Heidelberg (2011)
18. Heath, M., Bowyer, K., Kopans, D., Kegelmeyer Jr., W.P., Moore, R., Chang, K., Munishkumaran, S.: Current status of the digital database for screening mammography. In: Karssemeijer, N., Thijssen, M., Hendriks, J., van Erning, L. (eds.) Digital Mammography, pp. 457–460. Springer, Dordrecht (1998). doi:10.1007/978-94-011-5318-8_75
19. Badve, S., Bilgin, G., Dundar, M., Grcan, M.N., Jain, R.K., Raykar, V.C., Sertel, O.: Computerized classification of intraductal breast lesions using histopathological images. Biomed. Eng. **58**, 1977–1984 (2011). IEEE
20. Vedaldi, A., Matconvnet, L.K.: Convolutional neural networks for matlab. In: ACM International Conference on Multimedia 2015, pp. 689–692. ACM (2015)
21. Oliva, A., Torralba, A.: Modeling the shape of the scene: a holistic representation of the spatial envelope. IJCV **42**, 145–175 (2001)