Discriminative dictionary-embedded network for comprehensive vertebrae tumor diagnosis

Conference Paper · July 2020

CITATIONS
0

READS
16

5 authors, including:

Shen Zhao
Tsinghua University
32 PUBLICATIONS **127** CITATIONS
SEE PROFILE

Xi Wu
Chengdu University of Information Technology
93 PUBLICATIONS **435** CITATIONS
SEE PROFILE

Shuo Li
The University of Western Ontario
305 PUBLICATIONS **3,585** CITATIONS
SEE PROFILE

Discriminative dictionary-embedded network for comprehensive vertebrae tumor diagnosis

Shen Zhao[1,5], Bin Chen[2], Heyou Chang[3,5], Xi Wu[4], and Shuo Li[5]✉

[1] School of Intelligent Systems Engineering, Sun Yat-Sen University, Guangzhou, China, z-s-06@163.com
[2] ZheJiang University, ttbin@hotmail.com
[3] School of Information Engineering, Nanjing XiaoZhuang University, Nanjing, Jiangsu, China, cv_hychang@126.com
[4] Chengdu University of Information Technology, xi.wu@cuit.edu.cn
[5] Department of Medical Imaging and Medical Biophysics, Western University, London, Ontario, Canada, slishuo@gmail.com

Abstract. Comprehensive vertebrae tumor diagnosis (vertebrae recognition and vertebrae tumor diagnosis from MRI images) is crucial for tumor screening and preventing further metastasis. However, this task has not yet been attempted due to challenges caused by various tumor appearance, non-tumor diseases with similar appearance, irrelevant interference information, as well as diverse MRI image field of view (FOV) and/or characteristics. We purpose a **d**iscriminativ**e** di**ct**i**o**nary-embe**dd**ed n**e**twork (DECIDE) that contains an elaborated enhanced-supervision recognition network (ERN) and a discerning diagnosis network (DDN). Our ERN creatively designs projection-guided dictionary learning to leverage projections of angular point coordinates onto multiple observation axes for enhanced supervision and discriminability of different vertebrae. DDN integrates a novel label consistent dictionary learning layer into a classification network to obtain more discerning sparse codes for diagnosing performance improvement. DECIDE is trained and evaluated using a very challenging dataset consisted of 600 MRI images; the evaluation results show that DECIDE achieves high performance in both recognition (accuracy: 0.928) and diagnosis (AUC: 0.96) tasks.

Keywords: Vertebrae Tumor Diagnosis · Vertebrae Recognition · Dictionary Embedded Deep Learning.

1 Introduction

Comprehensive vertebrae tumor diagnosis (CVTD) means recognizing each vertebra by classifying its label and regressing its bounding box, and diagnosing whether it is invaded by tumor. CVTD is crucial because it joins recognition and diagnosis task together to enable direct diagnosis of vertebrae tumors, which are the most fatal spinal processes [1, 2], from magnetic resonance imaging (MRI)

images without manual processes such as vertebrae extraction. CVTD may clinically assist radiologists as an automated processor of MRI images for locating lesions, planning treatments, and preventing further metastasis [3, 4].

In CVTD, both recognition and diagnosis tasks are challenging. (1) The diagnosis task may be affected by various tumor appearances, e.g., local intensity changes in approximately circular or ring-like areas, global hypointense or hyperintense depending on MRI modalities, and diffuse pepper-salt like textures (Fig.1(a1∼a3)). Furthermore, it is difficult to distinguish from other spinal diseases such as end-plate osteochondritis (Fig.1(a4)). (2) The diagnosis task may suffer from massive irrelevant interference information in the non-vertebrae parts in the input MRI image. (3) The recognition task is troubled by unknown field of view (FOV) and corrupted MRI scans (e.g., under-sampled scans, severely deformed vertebrae) (Fig.1(a5)), which may lead to wrong-site surgery [5], a severe medical malpractice in clinical practice [6].

Very few work have been attempted for comprehensive vertebrae tumor diagnosis in the existing literature. [7] shows a closely relative work that detects tumors from MRI patches using convolutional neural networks (CNN). This work achieves high performance, however, it requires manual vertebrae extraction. [5, 8–10] singly perform vertebrae recognition, which lay solid foundation for further tumor diagnosis. [11, 12] use watershed algorithm and support vector machine for diagnosing spine metastases in CT images. However, this method cannot be applied in MRI images because intensity distributions and textures in MRI images are more complicated (Fig.1(a)). Other methods such as snakes, multi-task learning and state-space approaches can also be used in detection and/or segmentation tasks [13, 14]. In all, MRI is more sensitive for evaluating spinal lesions [7, 15], however, it is difficult for existing methods to distinguish vertebrae from non-vertebrae organs, and tumors from non-tumor tissues in corrupted or noisy MRI scans.

Dictionary learning (sparse coding) has the potential to obtain discriminative features from noisy images [16] for classification [17] and detection [18] tasks, however, two difficulties hinder its application in modern deep networks: (1) The dictionary and the sparse codes are generally trained in an alternate manner [19], which is difficult to be integrated into end-to-end training of CNN's. (2) The ground truth sparse codes are typically difficult to obtain. Traditional dictionary learning uses unsupervised reconstruction to obtain sparse codes, which may not be optimal for the main recognition [20] and diagnosis tasks [21].

We propose a novel discriminative dictionary-embedded network (DECIDE) for CVTD. DECIDE first recognizes each vertebra from input MRI, and then focuses on the features inside its bounding box for tumor diagnosis to eliminate interference information (tumor-like patterns of non-vertebrae parts). For both recognition and diagnosis tasks, DECIDE calculates sparse codes in the forward pass, which enables end-to-end dictionary training. Furthermore, in the recognition task, the angular points of the vertebrae's bounding boxes are sparse, i.e., they only account for very few proportions of pixels in the input image. This fact is leveraged to encode each vertebra by predicting L sparse codes via embedded

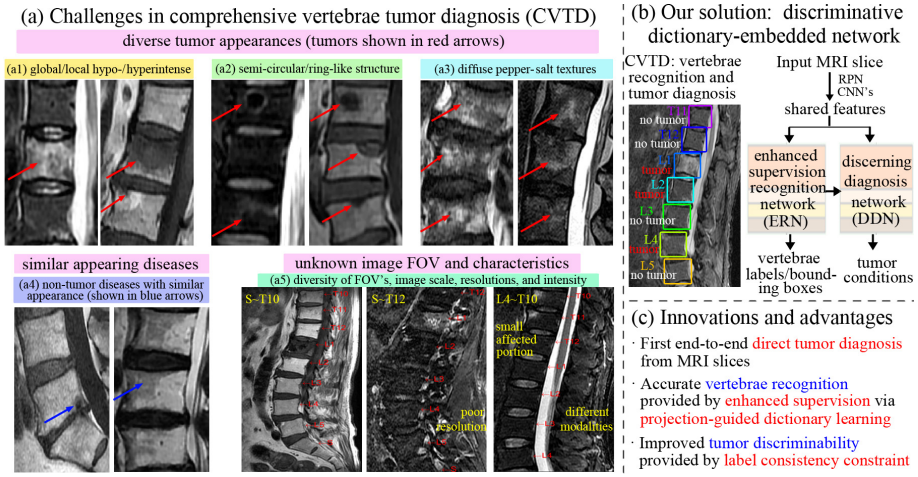

Fig. 1. Our proposed discriminative dictionary-embedded network (Fig.1(b)) addresses the challenges of comprehensive vertebrae tumor diagnosis (Fig.1(a), (a1∼a4) show various appearances of spinal tumors and similar appearances of non-tumor diseases, (a5) shows challenge of vertebrae recognition) by embedding dictionary into CNN-based recognition/classification networks Fig.1(c).

dictionary learning. The sparse codes are trained to approach the projections of ground truth angular points onto L observation axes (OA). For different vertebrae, since their projections on OA's of different orientations exhibit adequate discrepancy, the trained sparse codes have better distinguishability of different vertebrae. Under the projections' guidance, the ensemble of the predicted sparse codes helps to distinguish different vertebrae [22, 23]. In the diagnosis task, we diagnose tumor for each recognized vertebrae using features inside its bounding box to tackle interference information. A label consistent dictionary learning layer is designed to help minimize the Mahalanobis distance of vertebrae with the same diagnosis (cancer/non-cancer), which helps distinguish tumor from similar appearing diseases.

Our contributions are: (1) For the first time, vertebrae tumors are directly diagnosed from MRI images by an elaborated network that highlights key diagnostic information via vertebrae recognition. (2) An enhanced supervision method based on projection-guided dictionary learning is proposed for successfully recognizing sparsely distributed objects (vertebrae). (3) A label consistent strategy is introduced for improving classification (diagnosis) discriminability.

2 Methodology

Based on our pre-existing vertebrae recognition framework [5], we design our discriminative dictionary-embedded network (DECIDE, Fig.1(b)) by introducing the dictionary learning elements. DECIDE first adopts RPN [24] to coarsely

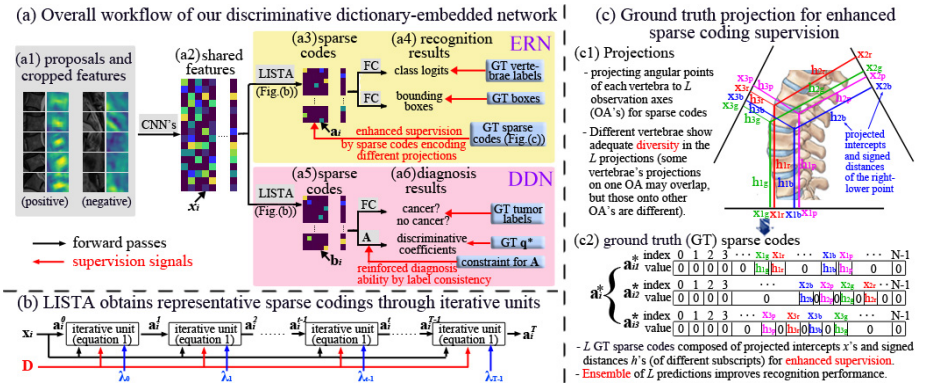

Fig. 2. Key modules of our approach. Fig.2(a) shows the overall workflow. Regional proposals and features are firstly obtained as in [24] (Fig.2(a1∼a2)). Our main contributions are: (1) The enhanced-supervision recognition network (ERN, in yellow background) that leverages the projections of the vertebrae's angular point coordinates to various observation axes (Fig.2(c)) for enhanced supervision. (2) The discerning diagnosis network (DDN, in pink background) that forces label consistency (i.e., vertebrae with the same diagnosis to have similar sparse codes) for enhanced diagnosis performance.

locate regions containing vertebrae. Then, two deliberate modules are designed: (1) Enhanced-supervision recognition network (ERN, Section 2.1, Fig.2, in yellow background) designs a feed-forward dictionary learning layer to obtain sparse codes. These sparse codes encode the projections of each vertebra on L observation axes for enhanced supervision, which helps improve the generalized vertebrae recognition accuracy and tackle the FOV/characteristics challenges. (2) Discerning diagnosis network (DDN, Section 2.2, Fig.2, in pink background) introduces a label consistent dictionary learning layer, which is embedded into a classification network to decrease the distance among sparse codes of vertebrae with the same diagnosis. This achieves accurate tumor diagnosis and alleviates the challenge caused by tumor appearances.

2.1 Enhanced-supervision recognition network (ERN)

Input of ERN. Our ERN takes the regional proposals (Fig.2(a1)) as its input. These proposals, obtained by regional proposal network (RPN) [24], are multi-scale rectangle boxes that coarsely cover vertebrae. For each proposal, its features are obtained by ROI aligning [24] and then flattened into vectors (denoted as $\mathbf{x}_i \in \mathbb{R}^M$ for the ith proposal) by cascading convolutional layers.

Acquiring representative sparse codes for recognizing different vertebrae. We design a dictionary learning layer to calculate the sparse code $\mathbf{a}_i$ for each $\mathbf{x}_i$. The subtlety of dictionary learning in recognition tasks is that the objects (vertebrae) are sparsely distributed, thus their positions (e.g., angular points) can be represented by $\mathbf{a}_i$ or its linear projection [18]. This draws forth the

idea of enhancing the supervision of sparse code for better $\mathbf{a}_i$, which resultantly yields better object locations. Meanwhile, it is confirmed in the compressive sensing community that $\mathbf{a}_i$ is able to be recovered by $\mathbf{x}_i$ (the output of CNN's) by minimizing $\frac{1}{2}\|\mathbf{x}_i - \mathbf{D}\mathbf{a}_i\|_2^2 + \lambda\|\mathbf{a}_i\|_1$ over $\mathbf{a}_i$. Inspired by the LISTA algorithm [25], we use Equ. 1 (visually demonstrated in Fig.2(b)) to obtain $\mathbf{a}_i$:

$$\mathbf{a}_i^t = \eta(\mathbf{a}_i^{t-1} + \beta\mathbf{D}^T(\mathbf{x}_i - \mathbf{D}\mathbf{a}_i^{t-1}); \lambda^t), \text{ where } \eta(\mathbf{r}; \lambda) = \text{sgn}(\mathbf{r}) \max\{|\mathbf{r}| - \lambda, 0\} \quad (1)$$

where the sparse code of the ith proposal $\mathbf{a}_i^t$ is updated iteratively by the shrinkage function η. η is a thresholding function that processes its input $\mathbf{r}$ element by element: For each element r_j, threshold λ is subtracted from its original absolute value $|r_j|$; if $|r_j| - \lambda < 0$, this element is set to 0 in the next iteration, which reduces non-zero elements. T iterations (typically $T = 3 \sim 6$) are applied to calculate reliable $\mathbf{a}_i$. In our design, the dictionary $\mathbf{D}$ as well as all λ^t's can be trained together with the preceding CNN's in an end-to-end manner.

Designing ground truth sparse codes. The key procedure of enhancing the supervision is to design appropriate ground truth sparse codes $\mathbf{a}_i^*$. For each vertebra (corresponding to a positive proposal), its ground truth sparse codes are obtained by the intercepts and signed distances of its angular point coordinates' projections to L observation axes around the input image, i.e., the projected intercepts x's and signed distances h's with different subscripts in $\mathbf{a}_i^*$ in Fig.2(c). The orientations of these axes are uniformly distributed (see Fig.2(c), for clarity, the projections of only one vertebra to $L = 3$ axes are demonstrated). For each non-vertebrae region (corresponding to a negative proposal), its ground truth sparse codes are set to zero vectors.

Loss function. After obtaining the predicted $\mathbf{a}_i$ and the ground truth $\mathbf{a}_i^*$, we design a loss function as follows: Firstly, two sibling fully connected (FC) layers are used as inverse projections; they take $\mathbf{a}_i$ as input and separately output L object class probability vectors and $4 \times L$ bounding box coordinates. Then, as in our previous work [5], message passing method is leveraged for vertebrae class probability calibration. Finally, all these calibrated class probabilities, bounding boxes, and sparse codes are supervised by the corresponding ground truths, i.e., the total loss function of the recognition task is:

$$\text{L}_\text{r} = \frac{\lambda_1}{N_1}\sum_{i_1=1}^{N_1}\sum_{l=1}^{L} L_{ce}(\mathbf{u}_{i_1,l}, u_{i_1}^*) + \frac{\lambda_2}{N_2}\sum_{i_2=1}^{N_2}\sum_{l=1}^{L} L_{sl}(\mathbf{v}_{i_2,l}, \mathbf{v}_{i_2}^*) + \frac{\lambda_3}{N_3}\sum_{i_3=1}^{N_3}\sum_{l=1}^{L} L_{sl}(\mathbf{a}_{i_3,l}, \mathbf{a}_{i_3,l}^*) \quad (2)$$

where: (1) $L_{ce}(\mathbf{u}_{i_1,l}, u_{i_1}^*)$ means the cross entropy loss of the predicted class probabilities $\mathbf{u}_{i_1,l}$ produced by the lth sparse code and the ground truth label $u_{i_1}^*$, N_1 is the total proposal number. (2) $L_{sl}(\mathbf{v}_{i_2,l}, \mathbf{v}_{i_2}^*)$ means the average of the smooth L1 loss [24] of all elements in vector $\mathbf{v}_{i_2,l} - \mathbf{v}_{i_2}^*$, i.e., the difference between the lth sparse code's prediction of the i_2th vertebra's bounding box coordinates $\mathbf{v}_{i_2,l}$ and the corresponding ground truth $\mathbf{v}_{i_2}^*$, N_2 is the positive proposal number. (3) $L_{sl}(\mathbf{a}_{i_3,l}, \mathbf{a}_{i_3,l}^*)$ means the smooth L1 loss of each predicted sparse code and its ground truth, N_3 is the total sparse code number.

Enhanced supervision. It is shown in Equ.2 that our ERN provides L supervision and L predictions (class probabilities, bounding boxes, and sparse codes) for each vertebra for enhanced supervision. Even if the projections of some vertebrae onto one axis overlap, those onto other axes can still show enough discrepancy because the OAs' orientations are diverse. This discrepancy helps distinguish different vertebrae. The ensemble of the L predictions determine the final recognitions, which helps lower risks of overfitting and better generalization performance [23].

Summarized advantages. Our ERN designs an end-to-end projection guided dictionary learning layer for enhanced supervision. The ensemble of a vertebra's L predicted projections onto different OA's improves the recognition discriminability and handles the FOV/image characteristics challenge.

2.2 Discerning diagnosis network (DDN)

Label consistent dictionary learning. Our DDN takes the features cropped by the recognized vertebrae as input, and then flattens them into vectors and solves for the sparse codes (denoted as $\mathbf{b}_i$ in diagnosis task for clarity) as mentioned above. However, since this task boils down to a classification problem, we do not have ground truth sparse codes for supervision. Thus, we impose a label consistency constraint to enhance discriminative capability by introducing the discriminative coefficient $\mathbf{q}_i \in \mathbb{R}^K$ for each vertebrae as follows:

Firstly, we uniformly assign labels to each element of $\mathbf{q}_i$; then, elements whose labels are the same with the ith proposal are assigned 1 and the others 0. In our work, $\mathbf{q}_i$ has K elements and 2 diagnosis labels (non-tumor and tumor), thus, the first $\frac{K}{2}$ elements in $\mathbf{q}_i$ are assigned label 0, and the last $\frac{K}{2}$ elements label 1; therefore, if the ith proposal has label 0, then the first $\frac{K}{2}$ elements in $\mathbf{q}_i$ are set to 1 and the others 0. $\mathbf{T}$ is the transformation matrix to predict sparse coefficients using $\mathbf{b}_i$. Thus, we have the label consistent loss term by calculating the Mahalanobis distance between the optimal sparse code $\mathbf{b}_i^*$ and predicted $\mathbf{b}_i$:

$$\|\mathbf{q}_i - \mathbf{T}\mathbf{b}_i\|_F^2 = \|\mathbf{T}\mathbf{b}_i^* - \mathbf{T}\mathbf{b}_i\|_F^2 = (\mathbf{b}_i^* - \mathbf{b}_i)^T \mathbf{T}^T\mathbf{T} (\mathbf{b}_i^* - \mathbf{b}_i) \tag{3}$$

In Equ. 3, $\mathbf{T}^T\mathbf{T}$ can be regarded as the covariance matrix in the definition of Mahalanobis distance. Although we do not explicitly know $\mathbf{b}_i^*$ in classification tasks, this term urges $\mathbf{b}_i$ to be close to it. Furthermore, since proposals with the same diagnosis labels have same $\mathbf{q}_i$'s, this term urges the Mahalanobis distance among their $\mathbf{b}_i$'s to be small. Since Mahalanobis distance has a strong discriminative capacity between data classes, the sparse codes are mapped to a more discriminative feature space than the input space. Thus, the label consistency dictionary learning mitigates diagnosis challenge [17].

Loss function. We formulate the label consistent loss term (Equ.3) into the total loss:

$$\mathrm{L_d} = \frac{\lambda_4}{N_2} \sum_{i=1}^{N_2} L_{ce}(\mathbf{c}_i, c_i^*) + \frac{\lambda_5}{N_2} \sum_{i=1}^{N_2} L_{sl}(\mathbf{q}_i, \mathbf{T}\mathbf{b}_i) \tag{4}$$

where: (1) $L_{ce}(\mathbf{c}_i, c_i^*)$ is the cross entropy classification loss of the predicted cancer logits $\mathbf{c}_i$ (which is obtained by simply feeding the sparse code $\mathbf{b}_i$ into a fully connected layer) and the ground truth diagnosis label c_i^*. N_2 is the positive proposal number as mentioned above. (2) $L_{sl}(\mathbf{q}_i, \mathbf{Tb}_i)$ is the label consistent loss term. We use the smooth L1 loss in Equ. 4 instead of the L2 loss in Equ. 3 to prevent exploding gradients when training together with the CNN's.

Summarized advantages. Our DDN purposes a label consistency constraint to prompt recognized vertebrae of the same diagnosis labels to be closer (i.e., the Mahalanobis distances among their sparse codes calculated using the features inside their bounding boxes are smaller). This tackles the interference information challenge and alleviates the tumor appearance challenge.

3 Experiments and Results

Dataset and ground truth annotations. A challenging dataset consisting of 600 spinal MRI images of ∼163 patients has been collected to demonstrate the effectiveness of our DECIDE for CVTD. The dataset contains arbitrary MRI images of thoracic, lumbar, and sacrum vertebrae of 6 different FOV's. For each patient, 3∼4 slices covering all vertebrae are chosen from 3D scans and resized to 512×512. Our dataset contains 4600 vertebrae, 818 of them are invaded by tumors. An experienced oncologist for spinal tumors has carefully labeled all vertebrae invaded by tumors twice with a temporal interval of one month. The second annotation is blinded to his initial annotation, which is used to assess intra-operator variability. The manual labels are used to set up the ground truth for training. If the two manual annotations have different labels (one indicates tumor and the other not), the ground truth is regarded as "positive" (invaded by tumors) for training.

Evaluation metrics. Standard five-fold cross-validation is used to evaluate our study. For the recognition performance, we follow the most recent vertebrae recognition work [5] using Image recognition accuracy (IRA), Identification rate (IDR), and mAP_{75} as evaluation metrics to respectively measure patient-wise accuracy, vertebrae-wise accuracy, and comprehensive classification-location performance. For the diagnosis performance, the ROC curve is used as the main evaluation metrics. Discussions based on this curve, such as area under curve (AUC), accuracy, precision, and recall are also conducted.

High recognition and diagnosis performances. Fig.3(a) shows that in MRI slices of various image FOV/characteristics and vertebrae appearance, the recognized vertebrae bounding boxes (dashed) overlaps well with the ground truth boxes (solid) of the correct labels (colors). Furthermore, the automatic diagnosis results (PRED) are generally the same as the ground truth (GT). In Fig.3(b), the black, red, and blue bars show high IRA (overall: 0.938, individual FOV's: >0.9), IDR (overall: 0.966, individual FOV's: >0.9), and mAP_{75} (overall: 0.956, individual FOV's: >0.92), which means that the recognition work produces very few wrongly classified, missing, or false positive recognitions. Fig.3(c) shows that our work achieves satisfactory tumor diagnosis ROC curve with an

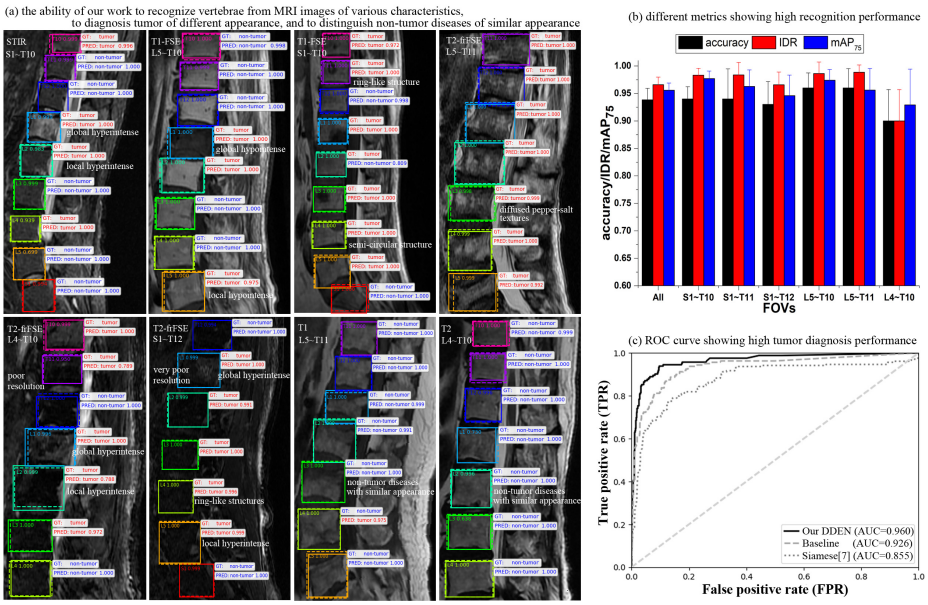

Fig. 3. Different metrics showing the effectiveness of our work. (a) Our network can accurately recognize vertebrae from images of different FOV and characteristics; it can also distinguish tumors from non-tumor diseases of various appearances. (b) Different metrics demonstrating high recognition performance. (c) The ROC curve and AUC value showing high diagnosis performance.

AUC of 0.96. If we regard vertebrae with predicted tumor probability greater than 0.5 as "positive" (suffering from tumors), we get a diagnostic accuracy of 0.939, precision of 0.860, and recall of 0.796. These results are comparable with those of the oncologist (second column in Table.1).

Table 1. Superiority of DECIDE to state-of-the-art methods and the ablation experiment without embedded dictionary.

Method	IRA	IDR	mAP_{75}	AUC	accuracy	precision	recall
Our DECIDE	0.928±0.021	0.956±0.014	0.946±0.013	0.960±0.073	0.939±0.102	0.860±0.168	0.796±0.183
intra-observer	–	–	–	–	0.964	0.874	0.838
Ablation	0.911±0.035	0.947±0.047	0.925±0.028	0.926±0.071	0.922±0.084	0.830±0.177	0.731±0.228
Siamese[7]	–	–	–	0.855±0.121	0.884±0.089	0.774±0.197	0.623±0.225
Hi-scene[5]	0.878±0.048	0.930±0.053	0.923±0.039	–	–	–	–
DI2IN[8]	0.803±0.149	0.904±0.115	–	–	–	–	–
Faster-RCNN[24]	0.750±0.138	0.869±0.104	0.848±0.146	–	–	–	–

Comparison with the state-of-the-art. Although no previous work has performed CVTD, we compare our work with those performing single recognition [5, 8, 24] or diagnosis [7] task. We also conduct ablation experiments using baseline methods without the dictionary learning layers. As shown in Table 1, DE-

CIDE outperforms all compared methods in both recognition (first three rows) and diagnosis (last four rows) tasks. For recognition, IRA, IDR, and mAP_{75} all benefit from the enhanced supervision provided by embedded dictionary. By comparing with the ablation study where the supervision by the projections on the L axes is disabled (Rows 1∼3 of the third column in Table 1), the advantage of enhanced supervision is demonstrated; even for the most difficult FOV T10∼L4, the performance is only slightly lower (Fig.3(b)). Furthermore, by comparing with [8] that uses dictionary learning as post-processing for landmark refinement, our embedded dictionary is more beneficial because it exploits the projections and introduces the ensemble of multiple predictions, which improves vertebrae recognition discriminability when their appearances are changed significantly by tumors. For diagnosis, the superiority of DECIDE to the ablation experiments (Rows 4∼7 of the third column in Table 1) shows that the label consistency strategy improves the diagnostic discriminability of sparse codes. Also, our DECIDE eliminates manual selection of patch size and resolution as in [7]; instead, the recognition network automatically deals with these issues and provides features adaptive to different vertebrae for succeeding diagnosis, i.e., the workflow of our DECIDE reinforces mutual benefits between two tasks and improves tumor diagnosis performance.

4 Conclusion

We have designed a discriminative dictionary-embedded network (DECIDE) as a novel clinical tool for comprehensive vertebrae tumor diagnosis (CVTD) from raw input MRI images. Our DECIDE integrates dictionary learning into both recognition and diagnosis networks for enhanced supervision and discriminative representations. The effectiveness of our network, as well as its advantage to the state-of-the-art, are demonstrated by extensive experiments.

References

1. Katherine N., Theresa A., Laurie K.: Cancer to bone: a fatal attraction. Nature Reviews Cancer, **11**(6), 411–425 (2011)
2. Mundy G.: Metastasis to bone: causes, consequences and therapeutic opportunities. Nature Reviews Cancer, **2**(8), 584–593 (2002)
3. OSullivan G., Carty F., Cronin C.: Imaging of bone metastasis: an update. World Journal of Radiology, **7**(8), 202–211 (2015)
4. Shelly S., Avi B., Orit S., Michal M., Hayit G., Eyal K.: Convolutional neural networks for radiologic images: A radiologists guide. Radiology, **290**(3), 590–606 (2019)
5. Zhao S., Wu X., Chen B., Li S.: Automatic vertebrae recognition from arbitrary spine MRI images by a hierarchical self-calibration detection framework. In: Shen D. et al. (eds.) Medical Image Computing and Computer Assisted Intervention - MICCAI 2019, Lecture Notes in Computer Science, vol. 11767, pp. 1–10. Springer, Cham (2019). https://doi.org/10.1007/978-3-030-32251-9_35
6. Philip F.: Patient safety in surgery. 2nd edn. Springer, Switzerland (2014)

7. Wang J., Fang Z., Lang N., Yuan H., Su M., Baldi P.: A multi-resolution approach for spinal metastasis detection using deep Siamese neural networks. Computers in Biology and Medicine, **84**(1), 137–146 (2014)
8. Yang D., Xiong T., Xu D., Huang Q., Liu D., Zhou S., Xu Z., Park J., Chen M., Tran T., Chin A., Metaxas D., Comaniciu D.: Automatic vertebra labeling in large-scale 3D CT using deep image-to-image network with message passing and sparsity regularization. In: Niethammer M. et al. (eds.) Information Processing in Medical Imaging – IPMI 2017, Lecture Notes in Computer Science, vol. 10265, pp. 633–644. Springer, Cham (2017). https://doi.org/10.1007/978-3-319-59050-9_50
9. Chen H., Shen C., Qin J., Ni D., Shi L., Cheng J., Heng P.: Automatic localization and identification of vertebrae in spine CT via a joint learning model with deep neural networks. In: Navab N., Hornegger J., Wells W., Frangi A. (eds.) Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015, Lecture Notes in Computer Science, vol. 9349, pp. 515–522. Springer, Cham (2015). https://doi.org/10.1007/978-3-319-24553-9_63
10. Lootus M., Kadir T., Zisserman A.: Vertebrae detection and labeling in lumbar MR images. In: Yao J., Klinder T., Li S. (eds.) Computational Methods and Clinical Applications for Spine Imaging. Lecture Notes in Computational Vision and Biomechanics, Lecture Notes in Computational Vision and Biomechanics, vol. 17, pp.219–230. Springer, Cham (2014)
11. Wiese T., Burns J., Yao J., Summers R.: Computer-aided detection of sclerotic bone metastases in the spine using watershed algorithm and support vector machines. In: 2011 IEEE International Symposium on Biomedical Imaging: From Nano to Macro, pp. 152–155. IEEE, Chicago, IL, USA, (2011). https://doi.org/10.1109/ISBI.2011.5872376
12. Burns J., Yao J., Wiese T., Munoz H., Jones E., Summers R.: Automated detection of sclerotic metastases in the thoracolumbar spine at CT. Radiology, **268**(1), 69–78 (2013)
13. Zhao S., Gao Z., Zhang H., Xie Y., Ghista D., Wei Z., Bi X., Xiong H., Xu C., Li S. Robust Segmentation of Intima-Media Borders with Different Morphologies and Dynamics During the Cardiac Cycle. IEEE Journal of Biomedical and Health Informatics, **22**(5), 1571–1582 (2018)
14. Gao Z., Li Y., Sun Y., Yang J., Xiong H., Zhang H., Liu X., Wu W., Liang D., Li S.. Robust estimation of carotid artery wall motion using the elasticity-based state-space approach. Medical image analysis, **37**, 1–21 (2017)
15. Shah L., Salzman K.: Imaging of spinal metastatic disease. International journal of surgical oncology, **2011**(1), 1–13 (2011)
16. Sun X., Nasrabadi N., and Tran T.: Supervised Deep Sparse Coding Networks for Image Classification. IEEE Transactions on Image Processing, **29**, 405–418 (2019)
17. Jiang Z., Lin Z., Davis L.: Label consistent K-SVD: Learning a discriminative dictionary for recognition. IEEE Transactions on Pattern Analysis and Machine Intelligence, **35**(11), 2651–2664 (2013)
18. Xue Y., Bigras G., Hugh J., Ray N.: Training Convolutional Neural Networks and Compressed Sensing End-to-End for Microscopy Cell Detection. IEEE Transactions on Medical Imaging, **38**(11), 2632–2641 (2019)
19. Aharon M., Elad M., Bruckstein A.: K-SVD: An algorithm for designing overcomplete dictionaries for sparse representation. IEEE Transactions on Image Processing, **54**(11), 4311–4322 (2006)
20. Zhao S., Wu X., Chen B., Li S.: Automatic spondylolisthesis grading from MRIs across modalities using faster adversarial recognition network. Medical image analysis, **58**, 101533 (2019)

21. Adam C. and Andrew N.: The importance of encoding versus training with sparse coding and vector quantization. Proceedings of the 28th international conference on machine learning (ICML2011), pp. 911–928. Omnipress, Bellevue, Washington, USA, (2011)
22. Xie H., Li J., and Xue H.: A survey of dimensionality reduction techniques based on random projection. arXiv preprint arXiv:1706.04371, (2017)
23. Quan Y., Xu Y., Sun Y., Huang Y., Ji H.: Sparse coding for classification via discrimination ensemble. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 5839–5847. IEEE, Las Vegas, NV, USA, (2016). https://doi.org/10.1109/CVPR.2016.629
24. Ren S, He K, Girshick R, Sun J.: Faster R-CNN: towards real-time object detection with region proposal networks. In: Advances in neural information processing systems (NIPS2015), pp. 91–99. Springer, Montreal, Quebec, Canada, (2015)
25. Gregor K, LeCun Y.: Learning fast approximations of sparse coding. In: Proceedings of the 27th International Conference on International Conference on Machine Learning (ICML2010), pp. 399–406. Omnipress, Madison, WI, USA (2010)