Using multi-task learning to improve diagnostic performance of convolutional neural networks

Mengjie Fang, Di Dong, Ruijia Sun, Li Fan, Yingshi Sun, et al.

Mengjie Fang, Di Dong, Ruijia Sun, Li Fan, Yingshi Sun, Shiyuan Liu, Jie Tian, "Using multi-task learning to improve diagnostic performance of convolutional neural networks," Proc. SPIE 10950, Medical Imaging 2019: Computer-Aided Diagnosis, 109501V (13 March 2019); doi: 10.1117/12.2512153

SPIE.

Event: SPIE Medical Imaging, 2019, San Diego, California, United States

Using Multi-task Learning to Improve Diagnostic Performance of Convolutional Neural Networks

Mengjie Fang[a,b], Di Dong[a,b], Ruijia Sun[c], Li Fan[d], Yingshi Sun[c], Shiyuan Liu[d], Jie Tian[a,b,*]
[a] CAS Key Laboratory of Molecular Imaging, Institute of Automation, Chinese Academy of Sciences, Beijing 100190, China
[b] University of Chinese Academy of Sciences, Beijing 100049, China
[c] Radiology Department, Peking University Cancer Hospital & Institute, Beijing 100142, China
[d] Department of Radiology, Changzheng Hospital, Second Military Medical University, Shanghai 200003, China

ABSTRACT

Due to the complex biological and physical mechanisms, the correlations between the classification objects of clinical tasks and the medical imaging phenotype are always ambiguous and implied, which makes it difficult to train a powerful diagnostic convolutional neural network (CNN) model efficiently. In this study, we propose a generic multi-task learning (MTL) CNN framework to achieve higher classification accuracy and better generalization. The proposed framework is designed to carry out the major diagnostic task and several auxiliary tasks simultaneously. It encourages the models to learn more beneficial representation following the underlying relation among patients' clinical characteristics, obvious imaging findings and quantitative imaging phenotype. We evaluate our approach on two clinical applications, namely advanced gastric cancer (AGC) serosa invasion diagnosis and discrimination of lung invasive adenocarcinoma manifesting as ground-glass nodule (GGN). Two datasets are utilized, which contain 357 AGC patients' venous phase contrast-enhanced CT volumes and 236 GGN patients' non-contrast CT volumes respectively. Several subjective CT morphology characteristics and common clinical characteristics are collected and used as the auxiliary tasks. To evaluate the generality of our strategy, CNNs with and without natural image-based pre-training are successively incorporated into the framework. The experimental results demonstrate that the proposed MTL CNN framework is able to improve the diagnostic performance significantly (7.4%-12.8% AUC increase and 3.5%-7.9% accuracy increase).

Keywords: Convolutional neural network, Multi-task learning, Computer-aided diagnosis, Computed tomography, Advanced gastric cancer, Lung cancer

1. INTRODUCTION

Deep learning methods, especially convolutional neural networks (CNNs), has improved the state-of-the-art in many medical applications ranging from tissue segmentation to cancer diagnosis.[1-3] Although so many advances have been achieved, the goal of incorporating these technologies into clinic procedures has not been accomplished. In regard to computer-aided diagnosis (CAD), the improvement of classification accuracy and the generalization of a system is always frustrated by the lack of training data, the noisy information in the images and the complex relationship between the clinical status and the medical imaging phenotype.[4] Therefore, more methods devoted to optimize the training procedure

* Corresponding author: Jie Tian, email: tian@ieee.org

Medical Imaging 2019: Computer-Aided Diagnosis, edited by Kensaku Mori, Horst K. Hahn, Proc. of SPIE Vol. 10950, 109501V · © 2019 SPIE
CCC code: 1605-7422/19/$18 · doi: 10.1117/12.2512153

and create links among multi-category characteristics of patients should be investigated.

In this study, we propose a generic framework utilizing the multi-task learning (MTL) to improve the diagnostic performance of CNN CAD models. Different from single-task learning (STL), MTL introduces inductive bias about the context to regularize the model.[5-6] The additional hypotheses could let the model learn more intrinsic relations between image and disease-related information while reduce the influence from other irrelevant noise patterns, as well as the risk of over-fitting. Meanwhile, taking the advantage of the relationship between the major task and the auxiliary tasks, the information learned from MTL could facilitate the classification. As a lot of clinical characteristics are usually hard to be collected, the proposed framework actually obtains a convenient approach to leverage the shared information across the variables. In particular, while the major diagnostic task and several auxiliary tasks simultaneously guide the training of the CNN modal through a weighted loss function, the inference about the auxiliary tasks are also used as references to the final classification.

2. METHODS

2.1 Multi-task Learning CNN Architecture

The proposed MTL CNN architecture is shown in Figure 1 and could be seen as an extension of the sing-task CNN architecture. The inserted CNN architecture is used as a quantitative feature extractor and yields a one-dimension feature vector of the input image. Then the feature vector (named FC0) is connected to two separated fully-connected layers (named FC1 and FC2). The layer FC1 is directly connected to several groups of softmax classifiers to predict some selected auxiliary characteristics, such as carcinoembryonic antigen (CEA) level, clinical T stage (reported by the radiologists) and subjective CT morphology. To leverage the reference information from the related tasks, the inferred nodes are then concatenated with the layer FC2 followed by a softmax classifier to carry out the major task.

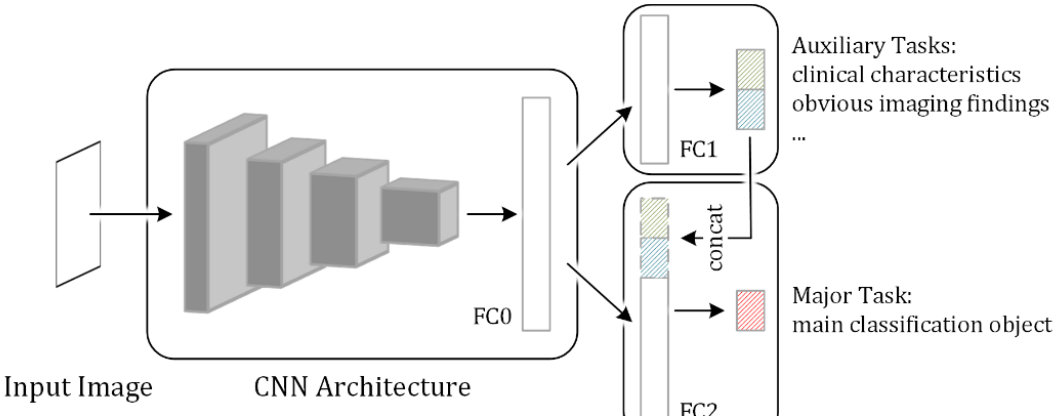

Figure 1. Illustration of MTL CNN architecture.

We define the loss function of the MTL model as a weighted sum of the losses of the major task L_{mt} and the auxiliary tasks L_{at} expecting that the CNN architecture could focus its attention on the intrinsic features about patient's clinical status. The loss function is formulated as:

$$L_{CNN} = m \times L_{mt} + (1 - m) \times L_{at} \quad (1)$$

where m is the combining weight and set to 0.7 in this study.

As the involved issues in CAD actually belong to the classification task, the cross-entropy is employed as the loss of the major task:

$$L_{mt} = -(1/N) \times \sum_{n=1}^{N} \sum_{c=1}^{C} g_c^n \log(p_c^n) \quad (2)$$

where N is the number of the images in a training batch, C is the number of the diagnosis categories, $g_c^n \in [0,1]$ and $p_c^n \in (0,1)$ are respectively the ground truth and the predicted result, i.e. the output of the softmax layer, corresponding to category c and image n.

Similarly, we use a linear weighted combination of cross-entropy to build the sum loss of the auxiliary tasks:

$$L_{at} = -(1/N) \times \sum_{n=1}^{N} \sum_{k=1}^{K} w_k \sum_{c=1}^{C_k} g_{k,c}^n \log\left(p_{k,c}^n\right) \quad (3)$$

where K is the number of auxiliary tasks, w_k is the weighting coefficient of the kth task and controls its importance in training.

To enhance the association between the feature representation and the major task, we set the weighting coefficient w_k of each auxiliary task to the Spearman's correlation coefficient between it and the major task. And we only involve the related tasks which show high statistical significance of p-values < 0.05 in Spearman's test. Noted that the ground truth values of the auxiliary characteristics are only needed in the training step.

2.2 Dataset

To investigate the performance of our MTL CNN framework in the diagnosis of common diseases, we compare it with the single-task CNN on the diagnosis of two clinical problems: 1. discriminating serosa invasion (i.e. T4 stage) in advanced gastric cancer (AGC); 2. discriminating invasive adenocarcinoma (IAC) in lung cancer manifesting as ground-glass nodules (GGN). This two problems both play important roles in determining the appropriate management of patients and the prediction of postoperative survival.[7-10]

For the first application, a dataset consisted of 357 AGC patients with 153 T3 and 204 T4 was collected from the Peking University Cancer Hospital & Institute. The venous phase CT volumes were acquired on a 64-slice CT scanner (LightSpeed 64; GE Healthcare). Clinical characteristics, such as CEA level, CA 199 level and CA 724 level, and subjective CT morphology characteristics interpreted by a radiologist, such as clinical T stage, nodular sign and cord sign, were collected and used as the classification objects of the auxiliary tasks.

For the second application, a dataset consisted of 236 GGN patients with 113 noninvasive lesion and 123 IAC was collected from the Changzheng Hospital. The non-contrast CT volumes were acquired on a 256-slice CT scanner (Brilliance-iCT; Philips Healthcare). Several kinds of subjective CT morphology characteristics, including lobulation, cusp angle and foam-like sign etc., were also collected.

In our experiments, the CT volumes are resampled to 0.5 mm/pixel in all three axes.

2.3 Implementation Details

In the experiment on each application, 60% and 10% of the patients are randomly selected and set aside as the training group and validation group respectively. The remaining 30% are used as the test group. For each patient's CT volume in training group, an experienced radiologist is asked to segment polygons around the tumor. Then we extract 100 64×64 patches randomly from the three planes (i.e. x-y, x-z and y-z) from each segmented region, and augment them 2 times with

random translations and rotations leading to around 64,000 and 42,000 training patches for the two application Meanwhile, for each volume in validation and test group, only one image patch located in the tumor center is marked and extracted. Noted that, in order to accommodate the AlexNet employed in this study, all extracted image patches are re-sampled via bi-linear interpolation to a fixed dimension of 224×224 pixels[11] and transformed into RGB channels via three CT window ranges[12]: [-800, -200 HU], [-200, 400 HU] and [-1000, 400 HU].

Considering that the small training size in medical diagnosis task can increase the difficulty of learning powerful feature representation and the over-fitting risk for the complex models, we employ the widely studied AlexNet[13] with convolutional layers in our experiment. Please note our framework can be easily extended to any other advanced CNN models. To train the network, we utilize two kinds of strategies: training it from scratch (random initialization, RI) and embedding the pre-trained parameter values into it followed by fine-tuning (transfer learning, TL). For the first strategy, initialize all parameters with random Gaussian distributions and train the network with mini-batch size of 50 patches Adam Optimizer with a learning rate of 0.001. We stop training when the network performance do not significantly improve on the validation set for 10 epochs. For the second strategy, we employ the convolutional architecture on the ImageNet dataset[14] and random initialize the fully-connected layers. The training setup of the second strategy is as same as that of the first one except that learning rate of fine-tuning the convolutional layers is set to 0.0002. We compare the proposed method with the default AlexNet (seen as a baseline single-task CNN model), of which softmax layer is changed to accommodate the major object categories, with the same two training strategies described above.

We report the area under the receiver operating characteristic curve (AUC), accuracy, sensitivity and specificity as metrics to assess the diagnostic performance of models.

3. RESULTS

The Spearman's correlation analysis reveals that there are six auxiliary characteristics relating with the application 1 and four auxiliary characteristics relating with the application 2, as listed in Table 1. These characteristics are used as the classification objects of the auxiliary tasks in our experiments.

Table 1. The selected auxiliary characteristics in our experiments.

Clinical application 1						
Characteristics	Clinical T stage	Clinical N stage	Nodular extension	Cord out layer	Fat infiltration	High enhanced serosa sign
Spearman's r	0.35	0.24	0.28	0.35	0.35	0.38
Clinical application 2						
Characteristics	Lobulation	Spiculation	Spine-like process	Air bronchogram		
Spearman's r	0.36	0.37	0.29	0.34		

Note. Spearman's r: Spearman's correlation coefficient

Table 2 summarizes the results of our approach and the compared STL architecture on two applications. The proposed MTL architecture with transfer learning achieves the highest AUC and accuracy. Noted that the MTL models outperform the STL models whether the pre-trained parameters are utilized or not. Leveraging the MTL architecture, the diagnostic performances of CNNs are improved obviously with 7.4%-12.8% AUC increase and 3.5%-7.9% accuracy

increase. This demonstrates that the assisting effect of the related tasks on training powerful CNNs is efficient and generalized.

The experimental result that the CNNs achieve more accurate diagnosis on the second clinical application indicates the common AlexNet is able to learn the significant feature representation for the GGN discrimination more easily. We infer the reason to be that lung lesions usually have obvious edge and strong CT textural phenotype which are similar with the characteristics of natural images. This is also confirmed by that transfer learning offered CNNs relatively larger performance improvement on application 2 (7.3% AUC increase) than application 1 (only 1.7% AUC increase). Oppositely, MTL CNNs outperform STL CNNs in all cases (more than 7.4% AUC increase). This implies that the proposed MTL architecture is able to facilitate the intrinsic feature extraction even when complex noise patterns exist.

Table 2. Comparison of different architectures for two clinical applications on the test sets.

Architecture	Clinical application 1				Clinical application 2			
	AUC	Accuracy	Sensitivity	Specificity	AUC	Accuracy	Sensitivity	Specificity
MTL + TL	**0.842**	**0.793**	**0.820**	0.756	**0.914**	**0.859**	**0.811**	0.912
MTL + RI	0.829	0.774	0.738	**0.822**	0.853	0.817	**0.811**	0.824
STL + TL	0.765	0.745	0.787	0.689	0.848	0.817	0.703	**0.941**
STL + RI	0.735	0.717	0.738	0.689	0.794	0.789	0.730	0.853

4. NEW OR BREAKTHROUGH WORK TO BE PRESENTED

There is lack of research on how to insert a CAD CNN model into MTL framework and how much it would help. We introduce a generic MTL CNN framework leveraging the auxiliary information from the related characteristics to achieve better performance. We evaluate the proposed MTL architecture on two clinical applications with transfer learning and with random initialization successively, and compare it with the baseline STL CNN. Numerical results indicate the potential of our approach to provide powerful support for improving the diagnostic performance of CNNs.

5. CONCLUSIONS

In this paper, we propose a MTL architecture to improve the diagnostic performance of CNNs. Taking advantage of the shared representations among related tasks, the proposed MTL architecture can encourage CNNs to extract more intrinsic and more significant features. The effectiveness and generalization of our method are testified through several comparative experiments. Future work will focus on optimizing the framework parameters and incorporating it with more skillful convolutional structures.

This work has not been submitted for publication or presentation elsewhere.

6. ACKNOWLEDGMENTS

This work was supported by the National Key R&D Program of China (2017YFA0205200, 2017YFC1308700, and 2017YFC1309100), National Natural Science Foundation of China (81771924, 81501616 and 81527805), the Beijing Natural Science Foundation (L182061), the Beijing Municipal Science and Technology Commission (Z171100000117023 and Z161100002616022), the Bureau of International Cooperation of Chinese Academy of Sciences (173211KYSB20160053), and the Youth Innovation Promotion Association CAS (2017175).

REFERENCES

[1] Van Tulder, G. and De Bruijne, M., "Combining generative and discriminative representation learning for lung CT analysis with convolutional restricted boltzmann machines," IEEE transactions on Medical Imaging 35(5), 1262-1272 (2016).

[2] Ronneberger, O., Fischer, P. and Brox, T., "U-net: Convolutional networks for biomedical image segmentation," in International Conference on Medical Image Computing and Computer-Assisted Intervention 234-241 (2015).

[3] Kooi, T., Litjens, G., van Ginneken, B., Gubern-Merida, A., Sánchez, C. I., Mann, R., den Heeten, A. and Karssemeijer, N., "Large scale deep learning for computer aided detection of mammographic lesions," Medical Image Analysis 35, 303-312 (2017).

[4] Shen, D., Wu, G. and Suk, H.-I., "Deep learning in medical image analysis," Annual Review of Biomedical Engineering 19, 221-248 (2017).

[5] Dai, J., He, K. and Sun, J., "Instance-aware semantic segmentation via multi-task network cascades," in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition 3150-3158 (2016).

[6] Twinanda, A. P., Shehata, S., Mutter, D., Marescaux, J., de Mathelin, M. and Padoy, N., "Endonet: A deep architecture for recognition tasks on laparoscopic videos," IEEE transactions on Medical Imaging 36(1), 86-97 (2017).

[7] Yoshikawa, T., Sasako, M., Yamamoto, S., Sano, T., Imamura, H., Fujitani, K., Oshita, H., Ito, S., Kawashima, Y. and Fukushima, N., "Phase II study of neoadjuvant chemotherapy and extended surgery for locally advanced gastric cancer," British Journal of Surgery 96(9), 1015-1022 (2009).

[8] Kunisaki, C., Akiyama, H., Nomura, M., Matsuda, G., Otsuka, Y., Ono, H.A., Nagahori, Y., Takahashi, M., Kito, F. and Shimada, H., "Surgical outcomes in patients with T4 gastric carcinoma," Journal of the American College of Surgeons 202(2), 223-230 (2006).

[9] Yanagawa, N., Shiono, S., Abiko, M., Ogata, S.Y., Sato, T. and Tamura, G., "New IASLC/ATS/ERS classification and invasive tumor size are predictive of disease recurrence in stage I lung adenocarcinoma," Journal of Thoracic Oncology 8(5), 612-618 (2013).

[10] Fan, L., Fang, M., Li, Z., Tu, W., Wang, S., Chen, W., Tian, J., Dong, D. and Liu, S., "Radiomics signature: A biomarker for the preoperative discrimination of lung invasive adenocarcinoma manifesting as a ground-glass nodule," European Radiology 29(2), 889-897 (2019).

[11] Anthimopoulos, M., Christodoulidis, S., Ebner, L., Christe, A. and Mougiakakou, S., "Lung pattern classification for interstitial lung diseases using a deep convolutional neural network," IEEE transactions on Medical Imaging 35(5), 1207-1216 (2016).

[12] Hoo-Chang, S., Roth, H.R., Gao, M., Lu, L., Xu, Z., Nogues, I., Yao, J., Mollura, D. and Summers, R.M., "Deep convolutional neural networks for computer-aided detection: CNN architectures, dataset characteristics and transfer learning." IEEE transactions on Medical Imaging 35(5), 1285-1298 (2016).

[13] Krizhevsky, A., Sutskever, I. and Hinton, G. E., "Imagenet classification with deep convolutional neural networks," in Proceedings of Advances in Neural Information Processing Systems 1097-1105 (2012).

[14] Deng, J., Dong, W., Socher, R., Li, L. J., Li, K. and Fei-Fei, L., "Imagenet: A large-scale hierarchical image database," in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition 248-255 (2009).