Unsupervised Contrastive Learning of Radiomics and Deep Features for Label-Efficient Tumor Classification

Ziteng Zhao[1] and Guanyu Yang[1,2(✉)]

[1] LIST, Key Laboratory of Computer Network and Information Integration (Southeast University), Ministry of Education, Nanjing, China
yang.list@seu.edu.cn

[2] Centre de Recherche en Information Biomédicale Sino-Français (CRIBs), Rennes, France

Abstract. Tumor classification is important for decision support of precision medicine. Computer-aided diagnosis by convolutional neural networks relies on a large amount of annotated dataset, which is costly sometimes. To solve the poor predictive ability caused by tumor heterogeneity and inadequate labeled image data, a self-supervised learning method combined with radiomics is proposed to learn rich visual representation about tumors without human supervision. A self-supervised pretext task, namely "Radiomics-Deep Feature Correspondence", is formulated to maximize agreement between radiomics view and deep learning view of the same sample in the latent space. The presented self-supervised model is evaluated on two public medical image datasets of thyroid nodule and kidney tumor and achieves high score on linear evaluations. Furthermore, fine-tuning the pre-trained network leads to a better score than the train-from-scratch models on the tumor classification task and shows label-efficient performance using small training datasets. This shows injecting radiomics prior knowledge about tumors into the representation space can build a more powerful self-supervised method.

Keywords: Self-supervised learning · Unsupervised contrastive learning · Radiomics · Tumor classification

1 Introduction

Deep convolutional neural networks (CNNs) have made major breakthroughs in the past few years, largely driven by increased computing power and massive labeled datasets. Benefitting from the huge advances of deep learning in image classification, computer-aided medical diagnostics has achieved great success [3]. Precise prediction of tumor type can help doctors recognize and interpret the subtle difference between different kinds of medical images. Moreover, it is critical to decision support of personalized cancer treatment for patients.

M. de Bruijne et al. (Eds.): MICCAI 2021, LNCS 12902, pp. 252–261, 2021.
https://doi.org/10.1007/978-3-030-87196-3_24

However, tumor classification using CNNs is still challenging due to (1) imaging data and tumor heterogeneity, and (2) the poor generalization ability of CNNs caused by inadequate labeled image data. Individual variability and the differences about acquisition protocols, contrast-agents, levels of contrast enhancements and scanner resolutions of medical image data lead to unpredictable size, shape and intensity diversity of tumors. Furthermore, up to date, most deep learning methods used in medical diagnosis are strongly supervised networks which require sufficiently large medical image datasets with expert annotations. Preparing large and labeled datasets is usually difficult or even impossible, with the result that CNNs cannot learn rich feature information and overfit severely.

To deal with complex data distribution and deficient annotated data, the self-supervised learning, a prominent pattern of unsupervised learning, is proposed to learn useful feature information from wide-ranging unlabeled data and then to solve the target task better with a small set of training data [1,12,13,15,18,21,23]. The key to self-supervised learning is to select suitable pretext tasks that can guide CNNs to extract high-quality visual features for the target tasks. Among many pretext tasks, unsupervised contrastive learning [1,8,23] is very popular in the field of natural images. It is a promising class of methods that build representations by learning to encode what makes two things similar or different. At a very high level, the intent of contrastive learning is to reduce feature dimensionality by maximizing agreement between different views of the same sample in the latent space. For example, CMC [19] learns invariant representations from various channels of one image; SimCLR [1] learns from differently augmented views of one image. For the applications of computer aided diagnosis, Jamaludin et al. [9] pre-trained a Siamese CNN distinguishing if a pair of images from different collection time is from the same patient. Jiao et al. [10] used cross-model contrastive learning to model the correspondence between video and audio of ultrasound. Therefore, it is evident that contrastive learning is a domain and task agnostic paradigm for self-supervised learning. It allows us to inject our prior knowledge about the structure in the data into the representation space and build more powerful self-supervised methods.

In tumor diagnosis, radiomics handcrafted quantitative features extracted from volumes of interest play an important role [4]. These features can describe a large number of phenotypic features, such as shape and texture [11]. From contrastive learning perspective, radiomics handcrafted features are a more effective view of medical images compared with the image augment transformation view because their feature dimensionality is low and they contain domain knowledge about diagnosis. Therefore, when contriving self-supervised pretext tasks, using contrastive learning of radiomics view and image view can help networks reduce feature dimensions and learn more discriminative features related to tumor diagnosis.

Therefore, we propose an unsupervised contrastive learning approach using radiomics and deep features for label-efficient tumor classification. We design a self-supervised pretext task, namely “Radiomics-Deep Feature Correspondence”,

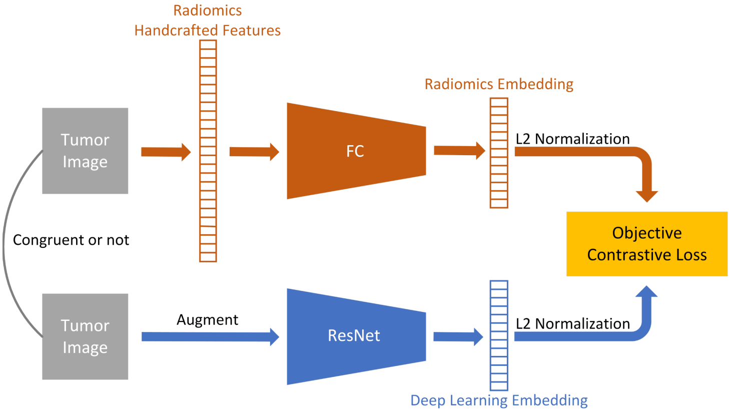

Fig. 1. Unsupervised contrastive learning by the pretext task of Radiomics-Deep Feature Correspondence. The pre-trained network is then fine-tuned to the target task.

to deeply exploit rich feature information from tumor areas and improve networks' ability to diagnose tumors in situations where labels are insufficient. We evaluate the presented method on two distinct public medical image datasets: thyroid nodule classification in ultrasound images and multiclass kidney tumor classification in CT. Experimental results show the proposed approach can learn better visual representation and its performance is superior to that of the network trained from scratch and is still good with less labeled training data. To the best of our knowledge, this is the first work achieving self-supervised learning by Radiomics-Deep Feature Correspondence pretext task.

2 Method

In this section, we begin with an overview of the self-supervised approach using Radiomics-Deep Feature Correspondence pretext task. We then introduce the details of the method, including contrastive loss of the self-supervised learning network and the full pipeline of the approach. The self-supervised learning network is illustrated in Fig. 1.

2.1 Radiomics-Deep Feature Correspondence

Our aim is to learn representations that hold information shared between radiomics view and deep learning view without human supervision. Radiomics handcrafted features extracted from tumor areas are the radiomics view's input and are processed by fully connected layers. Augmented Tumor images are the deep learning view's input and are processed by a CNN. After the above processing, we get radiomics features $r^{(n)}$ and deep learning features $d^{(n)}$ in the latent space and perform radiomics-deep feature correspondence pretext task by contrasting congruent and incongruent pairs. The self-supervised approach's label shows whether the two kinds of features, i.e., $r^{(n)}$ and $d^{(n)}$, are "corresponding". In other words, positive, $r^{(i)}$ and $d^{(i)}$ are from the same tumor area and should have similar distribution; negative, $r^{(i)}$ and $d^{(j)}$ are from different tumor areas. The pretext task can help networks learn a representation that maximize mutual information between the two views of the same tumor but is otherwise compact.

2.2 Contrastive Loss and Memory Bank

To achieve the pretext task's aim, we apply contrastive learning [2,23], where feature embeddings such that two views of the same tumor map to nearby points while views of different tumor to far apart points. The distance of points is measured with cosine similarity in representation space. In practice, we train InfoNCE [5,23] loss to correctly select d corresponding to r out of a set $S = \{r, r^{(1)}, r^{(2)}, \ldots, r^{(k)}\}$ that contains k incongruent radiomics features, i.e., select a only positive sample from a set that contains k negative samples. This loss function $\mathcal{L}^{d,r}_{contrast}$ is shown below.

$$\mathcal{L}^{d,r}_{contrast} = -log \frac{exp(d^\top r)}{\sum_{i \in k} exp(d^\top r^i)} \tag{1}$$

Loss $\mathcal{L}^{d,r}_{contrast}$ in Eq. 1 treats deep learning view as anchor and enumerates over radiomics view. Symmetrically, we can get $\mathcal{L}^{r,d}_{contrast}$ by anchoring at radiomics view. We add the two loss up as the final loss:

$$\mathcal{L}(r, d) = \mathcal{L}^{d,r}_{contrast} + \mathcal{L}^{r,d}_{contrast} \tag{2}$$

Better representations using $\mathcal{L}(r, d)$ in Eq. 2 can be learnt by using lots of negative samples [17]. In order to reduce the computational cost of calculating a large number of negative samples, we maintain two memory banks to store two views' feature embeddings for each training sample following [23]. The memory banks are dynamically updated with latent features computed on the fly.

2.3 The Full Framework of the Approach

We first pre-train the network in Fig. 1 on radiomics-deep feature correspondence pretext task. Upon convergence, we add a fully-connected layer at the end of the pre-trained network and use it for tumor classification.

Radiomics Features Extraction. Around 1,000 radiomics handcrafted features are extracted from each tumor area by PyRadiomics [20]. PyRadiomics is a comprehensive open-source python package, which is able to extract reproducible handcrafted features through a large panel of hard-coded feature algorithms. These features which contain intensity-based features, shape-based features, texture-based features and higher-order features are based on domain knowledge and can characterize tumor heterogeneity to some extent. And then they are standardized and processed by three fully connected layers to 128-dimensional feature vectors.

Deep Learning Features Extraction. 2D ResNet-50 [6,7] is used to extract deep features from each tumor area. For the input of CNN, different sized tumor areas are resized to the average size and then normalized to $[0, 1]$. In addition, we apply an augmentation which is random cropping followed by resize back to the original size. In terms of network architecture, the backbone network can be any well-structured 2D CNN. Here we choose ResNet empirically, which is one of the best performing and generic networks. For the network's output, deep learning features are generated by global average pooling of the last convolutional layers of ResNet. And they are converted to 128-dimensional feature vectors, which are consistent with the dimension of radiomics features. Additionally, we constrain two views's embeddings by L_2 normalization before calculating the contrastive loss function, as suggested in [23].

3 Experiment

This section first introduces the two public medical image datasets and implementation details about experiments. We then evaluate our method using linear classification and transfer learning, and demonstrate that it's an effective self-supervised mechanism to classify tumor types.

3.1 Dataset

Thyroid Nodule Classification in Ultrasound Images. The challenge of Thyroid Nodule Segmentation and Classification in Ultrasound Images (TN-SCUI2020) [14] provide a public 2D dataset of thyroid nodule with over 3,644 patient cases from different ages, genders, and were collected in different sites using various ultrasound machines (e.g. Mindray DC-8, Philips-cx50, TOSHIBA Aplio300). Each ultrasound image is provided with its annotated class (benign or malignant) and a detailed delineation of the nodule. For pre-processing, we crop nodule areas from images and resize the areas to 196 × 160 in dimension. The dataset is randomly split to a train set (2,916 cases) and a test set (728 cases) by the ratio of 80 : 20, while preserving the percentage of samples for each class. In self-supervised learning, the train set is used for the pretext task.

Multiclass Kidney Tumor Classification in CT. The challenge of 2019 Kidney and Kidney Tumor Segmentation (KiTS19) [22] released 210 3D abdominal CT images with kidney tumor subtypes and segmentations of kidney and kidney tumor. These CT images are from more than 50 institutions and scaned with different CT scanners and acquisition protocols. There are many subtypes of tumor in the dataset: clear cell renal cell carcinoma (RCC) (143 cases), papillary RCC (21 cases), chromophobe RCC (19 cases), oncocytoma (10 cases) and other smaller classes. We classify kidney tumors into four larger categories. In order to balance the quantity in each category, we randomly select 20 cases from the clear cell RCC category and combine them with the other three types of data to form a classification data set (70 cases). The remaining data (140 cases) in the KiTS19 dataset is used for self-supervised learning. For pre-processing, we uniformly select 10 slice images for each tumor area because of a wide range of slice thicknesses. If there are less than 10 slices, select slices repeatedly from the middle to the ends. And we resize these tumor slices to 10 × 64 × 64, and then send them to CNN as a whole. For classification, we perform patient wise five-fold cross-validation.

3.2 Implementation Details

Our approach is implemented using PyTorch 1.7.1 and trained with a single NVIDIA GeForce GTX 2080Ti. And all the networks are optimized with stochastic gradient descent (SGD). In the self-supervised training phase, the model is trained up to 700 epochs with a learning rate of 3e–3. The capacity of the memory bank is the size of the entire training data set, the temperature is 0.07 and a momentum for memory update is 0.5. In the tumor classification training phase, all the models are trained for 100 epochs with a learning rate of 3e–3 and decayed by a factor of 10 at the 70th epoch. We use accuracy and weighted F1 score [16] to evaluate the models' performance.

Table 1. Linear classification performance of self-supervised pretext tasks. The table shows that the representation obtained by our self-supervised method performs superior on the given datasets.

Dataset	Self-supervised method	F1 score (weighted) (%)
Thyroid nodule	Autoencoder [21]	68.9
	Jigsaw [15]	71.6
	SimCLR [1]	78.7
	Ours	82.9
Kidney tumor	Autoencoder [21]	46.4 ± 5.3
	Jigsaw puzzles [15]	45.0 ± 3.5
	SimCLR [1]	50.0 ± 4.5
	Ours	52.0 ± 2.8

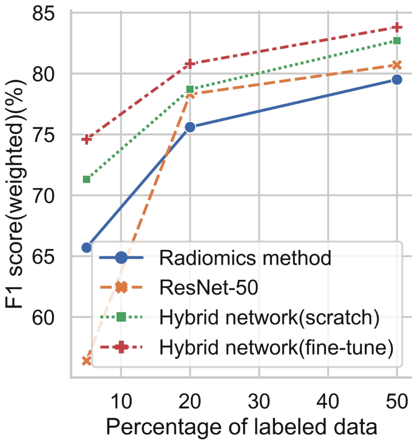

Fig. 2. Label-efficient image classification with the pre-trained model by radiomics-deep feature correspondence self-supervised learning.

Table 2. Evaluations of tumor classification using different models. The table shows that radiomics handcrafted features are useful in tumor diagnosis and fine-tuning our self-supervised model can improve the performance.

Dataset	Method	Accuracy (%)	F1 score (weighted) (%)
Thyroid nodule	Radiomics method [4]	81.3	80.7
	ResNet-50 [6]	82.6	82.2
	Hybrid network (scratch)	83.4	83.1
	Hybrid network (fine-tune)	84.9	84.4
Kidney tumor	Radiomics method [4]	56.2 ± 6.4	56.8 ± 7.0
	ResNet-50 [6]	50.0 ± 5.3	49.3 ± 7.6
	Hybrid network (scratch)	52.8 ± 5.7	52.0 ± 5.6
	Hybrid network (fine-tune)	64.3 ± 5.9	63.7 ± 6.9

3.3 Linear Classification on Self-supervised Model

Linear Classification is a general benchmark to evaluate unsupervised image representations' quality. We fix the pre-trained network as shown in Fig. 1 and add a fully connected layer behind the generated embeddings for evaluation. In addition, we compare other self-supervised methods using ResNet50. As Table 1 shows, F1 score of radiomics-deep feature correspondence network is higher than models that perform other pretext tasks. This indicates that the self-supervised task which incorporates prior knowledge can improve model's representations for tumors. Furthermore, the linear evaluation results of our model are comparable to the results of supervised models in Table 2. This implies that the embeddings obtained by our pretext task are discriminative and useful representations of tumors, although the distribution of medical image data is complicated.

3.4 Benefits of Self-supervised Pre-training

We now investigate the question of whether radiomics-deep feature correspondence method can improve the performance of image classification, even in the situation of small datasets, i.e., the kidney tumor dataset and a small part of the thyroid nodule dataset.

Supervised Baseline. We conduct three benchmark experiments. First, we use a conventional radiomics method [4] to analyze radiomics handcrafted features, that is, the pipeline of feature preprocessing, feature selecting, and classification. Second, we train ResNet-50 from scratch. Third, we train a hybrid network that combines radiomics and deep learning, namely the network in Sect. 3.3. But all the parameters in this model are random and all updated. The results are shown in Table 2. We can see that the radiomics method perform well on tumor classification, so some of the radiomics handcrafted features can indeed be correlated with tumor label. Moreover, the hybrid model that incorporates prior knowledge can improve performance compared to pure ResNet-50. In summary, quantitative radiomics handcrafted features can reflect the heterogeneity of tumors and increase the power of deep learning models for classification.

Transfer Learning Using All Labeled Data. We fine-tune the pre-trained hybrid network using our self-supervised task. We experimented with fine-tuning different numbers of layers. And we found that the best results were obtained without fine-tuning the last group of convolutions layers in ResNet-50 and the last two layers in fully connected network for processing radiomics handcrafted features. As Table 2 shows, evaluation scores for fine-tuning the pre-trained hybrid network are higher than training from scratch. It is noteworthy that fine-tuning the pre-trained hybrid network can achieve high scores in the kidney tumor classification task with less data and more types. This demonstrates the self-supervised approach can learn effective visual representations for tumor classification and the pre-trained model can generate more discriminative features for the target task by fine tuning.

Efficient Learning Using Less Labeled Data. We evaluate the performance of fine-tuning the pre-trained hybrid model as the size of the labeled dataset varies to 5%, 20%, 50%. The test dataset is always the same as the default setting. We only train these experiments on the thyroid nodule dataset, because the kidney tumor dataset is too small to do that. For comparison, the radiomics method, randomly-initialized ResNet-50 and hybrid network are trained with the above settings. The weighted F1 scores from models with different training datasets are displayed in Fig. 2. Compared to the three train-from-scratch models, the network with pre-trained weights always maintains high performance and brings a more remarkable gain with decreasing amounts of labeled data. These results reveal that our approach alleviates the current situation of insufficient medical labeled data to some extent. By the radiomics-deep feature correspondence pre-training method, the network learns rich visual features from a large amount of unlabeled data and can get free performance improvements with zero manual annotations.

4 Conclusion

In this paper, the radiomics-deep feature correspondence self-supervised learning approach is proposed to boost the model's predictive power of the tumor types. Contrasted by the radiomics view which contains phenotypic characteristics and domain knowledge, the hybrid network can learn good visual representation and mitigate overfitting when solving the target task with insufficient labeled data. Our method is validated on two public medical image datasets to demonstrate its label-efficient classification performance. Future works include enhancement of pre-trained models with more data in public medical image datasets and applications of pre-trained models used in other medical tasks.

Acknowledgements. This research was supported by National Natural Science Foundation under grants (31571001, 61828101). We thank the Big Data Center of Southeast University for providing the GPUs to support the numerical calculations in this paper.

References

1. Chen, T., Kornblith, S., Norouzi, M., Hinton, G.: A simple framework for contrastive learning of visual representations. In: International Conference on Machine Learning, pp. 1597–1607. PMLR (2020)
2. Chopra, S., Hadsell, R., LeCun, Y.: Learning a similarity metric discriminatively, with application to face verification. In: 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05), vol. 1, pp. 539–546. IEEE (2005)
3. Esteva, A., et al.: A guide to deep learning in healthcare. Nat. Med. **25**(1), 24–29 (2019)
4. Gillies, R.J., Kinahan, P.E., Hricak, H.: Radiomics: images are more than pictures, they are data. Radiology **278**(2), 563–577 (2016)
5. Gutmann, M., Hyvärinen, A.: Noise-contrastive estimation: a new estimation principle for unnormalized statistical models. In: Proceedings of the Thirteenth International Conference on Artificial Intelligence and Statistics, pp. 297–304. JMLR Workshop and Conference Proceedings (2010)
6. He, K., Zhang, X., Ren, S., Sun, J.: Deep residual learning for image recognition. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 770–778 (2016)
7. He, T., Zhang, Z., Zhang, H., Zhang, Z., Xie, J., Li, M.: Bag of tricks for image classification with convolutional neural networks. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 558–567 (2019)
8. Henaff, O.: Data-efficient image recognition with contrastive predictive coding. In: International Conference on Machine Learning, pp. 4182–4192. PMLR (2020)
9. Jamaludin, A., Kadir, T., Zisserman, A.: Self-supervised learning for spinal MRIs. In: Cardoso, M. et al. (eds.) Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support. LNCS, vol. 10553, pp. 294–302. Springer, Cham (2017). https://doi.org/10.1007/978-3-319-67558-9_34
10. Jiao, J., Cai, Y., Alsharid, M., Drukker, L., Papageorghiou, A.T., Noble, J.A.: Self-supervised contrastive video-speech representation learning for ultrasound. In: Martel, A.L. et al. (eds.) International Conference on Medical Image Computing and Computer-Assisted Intervention. MICCAI 2020. LNCS, vol. 12263, pp. 534–543. Springer, Cham (2020). https://doi.org/10.1007/978-3-030-59716-0_51
11. Lambin, P., et al.: Radiomics: the bridge between medical imaging and personalized medicine. Nat. Rev. Clin. Oncol. **14**(12), 749 (2017)
12. Larsson, G., Maire, M., Shakhnarovich, G.: Colorization as a proxy task for visual understanding. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 6874–6883 (2017)
13. Nathan Mundhenk, T., Ho, D., Chen, B.Y.: Improvements to context based self-supervised learning. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 9339–9348 (2018)
14. Ni, D.: Thyroid nodule segmentation and classification in ultrasound images (tn-scui2020) (2020). https://tn-scui2020.grand-challenge.org/

15. Noroozi, M., Favaro, P.: Unsupervised learning of visual representations by solving Jigsaw puzzles. In: Leibe, B., Matas, J., Sebe, N., Welling, M. (eds.) European Conference on Computer Vision. ECCV 2016. LNCS, vol. 9910, pp. 69–84. Springer, Cham (2016). https://doi.org/10.1007/978-3-319-46466-4_5
16. Pedregosa, F., et al.: Scikit-learn: machine learning in Python. J. Mach. Learn. Res. **12**, 2825–2830 (2011)
17. Poole, B., Ozair, S., Van Den Oord, A., Alemi, A., Tucker, G.: On variational bounds of mutual information. In: International Conference on Machine Learning, pp. 5171–5180. PMLR (2019)
18. Tao, X., Li, Y., Zhou, W., Ma, K., Zheng, Y.: Revisiting Rubik's cube: self-supervised learning with volume-wise transformation for 3D medical image segmentation. In: Martel, A.L. et al. (eds.) International Conference on Medical Image Computing and Computer-Assisted Intervention. MICCAI 2020. LNCS, vol. 12264, pp. 238–248. Springer, Cham (2020). https://doi.org/10.1007/978-3-030-59719-1_24
19. Tian, Y., Krishnan, D., Isola, P.: Contrastive multiview coding. In: Vedaldi, A., Bischof, H., Brox, T., Frahm, J.M. (eds.) Computer Vision – ECCV 2020. LNCS, vol. 12356, pp. 776–794. Springer, Cham (2020). https://doi.org/10.1007/978-3-030-58621-8_45
20. Van Griethuysen, J.J., et al.: Computational radiomics system to decode the radiographic phenotype. Cancer Res. **77**(21), e104–e107 (2017)
21. Vincent, P., Larochelle, H., Bengio, Y., Manzagol, P.A.: Extracting and composing robust features with denoising autoencoders. In: Proceedings of the 25th International Conference on Machine Learning, pp. 1096–1103 (2008)
22. Weight, C.: The 2019 kidney and kidney tumor segmentation challenge (KiTS19) (2019). https://kits19.grand-challenge.org/
23. Wu, Z., Xiong, Y., Yu, S.X., Lin, D.: Unsupervised feature learning via non-parametric instance discrimination. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 3733–3742 (2018)