Collage CNN for Renal Cell Carcinoma Detection from CT

Mohammad Arafat Hussain[1(✉)], Alborz Amir-Khalili[1], Ghassan Hamarneh[2], and Rafeef Abugharbieh[1]

[1] BiSICL, University of British Columbia, Vancouver, BC, Canada
{arafat,alborza,rafeef}@ece.ubc.ca
[2] Medical Image Analysis Lab, Simon Fraser University, Burnaby, BC, Canada
hamarneh@sfu.ca

Abstract. Renal cell carcinoma (RCC) is a common malignancy that accounts for a steadily increasing mortality rate worldwide. Widespread use of abdominal imaging in recent years, mainly CT and MRI, has significantly increased the detection rates of such cancers. However, detection still relies on a laborious manual process based on visual inspection of 2D image slices. In this paper, we propose an image collage based deep convolutional neural network (CNN) approach for automatic detection of pathological kidneys containing RCC. Our collage approach overcomes the absence of slice-wise training labels, enables slice-reshuffling based data augmentation, and offers favourable training time and performance compared to 3D CNNs. When validated on clinical CT datasets of 160 patients from the TCIA database, our method classified RCC cases vs. normal kidneys with 98% accuracy.

1 Introduction

Renal cell carcinomas (RCC) refer to a group of chemotherapy-resistant malignancies [1] that constitutes the most common type of kidney cancer in adults. Responsible for approximately 90% of cases of kidney cancers, RCCs accounted for an estimated 61,560 new patients and 14,080 deaths in the United States in 2015 alone [2]. In fact, North America and Europe have recently reported the highest numbers of new cases of renal tumor in the world [3]. This increased number of RCC over the past several years has been contributing to a steadily increasing mortality rate per unit population worldwide [4].

Although some patients with renal tumors present with clinical symptoms like flank pain, gross haematuria or palpable abdominal mass, the detection rate of renal tumours has significantly increased due to the widespread use of various types of abdominal imaging including ultrasonography, computed tomography (CT) and magnetic resonance imaging (MRI). Typically, tumour staging is accomplished with CT, which allows for assessment of local invasiveness, lymph node involvement, or other metastases. Nonetheless, more than 50% of RCCs are currently detected incidentally [4]. This RCC detection is typically carried out by radiologists through manual observation of abdominal image data. Although a

Q. Wang et al. (Eds.): MLMI 2017, LNCS 10541, pp. 229–237, 2017.
DOI: 10.1007/978-3-319-67389-9_27

good number of studies have been carried out on kidney localization and anatomical analysis [5–9], to the best of our knowledge, there has been no study to date that focused on automatic discrimination between healthy vs. renal cell carcinoma kidneys. Such discrimination ability coupled with an automatic kidney localization procedure would be invaluable during targeted as well as incidental analysis of kidney health.

Medical image analysis has enjoyed significant performance improvements through the use of various machine learning (ML) algorithms over the past few years. Most of these algorithms are fully supervised, requiring a large number of annotated datasets for model learning and prediction accuracy analysis. Unlike two dimensional (2D) single- or three-channel data (e.g., gray-scale or color images), which are most commonly used in computer vision tasks, three dimensional (3D) medical data presents different sets of challenges for ML approaches. For example, tissue abnormalities such as tumors, cancers, nodules, stones etc. are most often localized within a small region of anatomy and do not span the whole image volume. Localization and analysis of abnormal tissue are thus typically carried out on the 2D image slices. For example, staging of kidney tumors is done through slice-based tumor analysis and manual boundary tracing. However, image tags or labels (e.g. healthy, cancerous etc.) are mostly assigned per image volume or per patient basis. Therefore, all slices of an image are by default labeled with a single tag, though not all slices may contain the abnormal tissue. This scenario makes 'single-instance' ML approaches, especially deep learning ones such as convolutional neural networks (CNNs), very difficult to train on the 2D slices, as the input slice often does not correspond to the assigned volume-based label. A typical solution for this problem is to use the full 3D image volume as a single-instance for learning. However, 3D CNNs are considerably more difficult to train as they contain significantly more parameters and consequently require many more training samples, necessitate the use of expensive GPUs with very large memory, and require a lot more time to converge.

An alternative approach to single-instance learning is multiple-instance learning (MIL) [10]. MIL is a variation on weakly supervised learning wherein the learner receives a set of labeled bags, or ensembles, each containing multiple instances. This scenario allows the learner to label a bag with a class even if some or most of the instances within it are not members of that class. Using this MIL approach, the objective of our RCC detection application can be formulated such that a labelled bag corresponds to a labeled CT volume, and the constituting instances within the bag correspond to the CT's 2D slices, some of which may contain RCC tumors while many may not. This reformulation allows us to correctly incorporate volume-based labels within an easy to train 2D slice-based CNN framework. In the context of deep learning on medical images, the joint benefits of MIL combined with the classification power of 2D CNNs have been recently demonstrated in a few applications including mammogram classification for breast cancer detection [11], identifying anatomical body parts [12], colon cancer classification based on histopathology images [13], and classification of large 2D microscopy images [14]. To the best of our knowledge, such

an approach has not been implemented specifically on 3D kidney data and a novel representation of volumetric CT data is necessary in order to extend such techniques for detection of RCC in CT data.

In this paper, we propose a CNN based kidney classification method that makes use of a novel collage image representation. The image slices in a 3D volume are rearranged side-by-side into a virtual extended 2D image slice, which in turn correctly corresponds to the single available label for that dataset. Our approach is different from Zhu et al. [11] and Kraus et al. [14] as, instead of explicitly modelling MIL aggregation as a global pooling layer, we design the architecture of our CNN to implicitly learn a nonlinear relationship between the bag labels and feature representation of the encapsulated instances. Compared to the computationally expensive two stage (i.e. pre-train and boosting stages) CNN learning procedure adopted by Yan et al. [12], our single collage CNN is trained end-to-end and effectively performs efficient classification. Xu et al. [13] took the advantage of a fully supervised classifier (support vector machines) on the slice-wise labeled data along with an MIL procedure on the rest of the weekly labeled data. Our proposed collage also allows for data augmentation by random reshuffling of the locations of axial image slices within the collage; this augmentation also facilitates the training of the implicit relationship between bag labels and learn feature representation.

2 Materials and Methods

2.1 Data

Our clinical dataset consisted of 160 kidney scans of 160 patients accessed from The Cancer Imaging Archive (TCIA) database [15]. We used 80 healthy kidney samples from 80 patients who had one healthy kidney. The 80 pathological kidney samples used were from another 80 patient scans. Our dataset had variations in the scanner types, contrast administration, fields of view, spatial resolutions, and intensity (Hounsfield unit) ranges. The in-plane pixel size ranged from 0.58 to 1.50 mm and the slice thickness ranged from 1.5 to 5 mm. Of the 80 healthy and 80 carcinoma scans, we randomly chose 55 cases from each set to use for training and the remaining 25 for testing. Ground truth kidney RCC labels were also collected from the TCIA data records.

2.2 Collage Representation of 3D Image Data

Typically, renal tumors grow in different regions of the kidney and are clinically scored on the basis of their CT slice-based image features such as size, margin (well-define or ill-defined), composition (solid or cystic), necrosis, growth pattern (endophytic or exophytic), calcification etc. [16]. Of course not all kidney slices necessarily contain tumors, nonetheless clinical labels (healthy/pathological) are normally recorded on a kidney- or a patient-basis. Therefore, it is not possible to use slice-based inputs in the training of a CNN because the volume-based label is

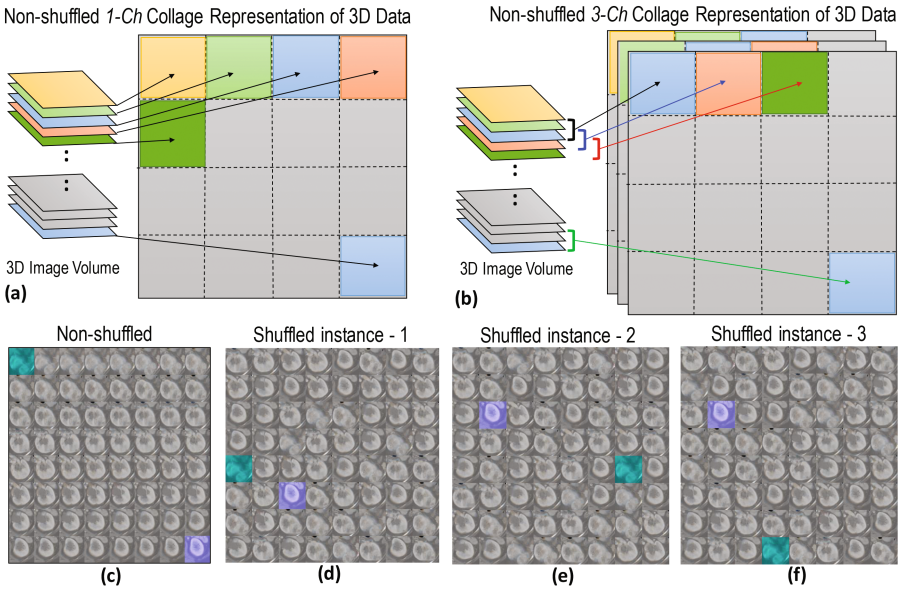

Fig. 1. Schematic diagrams showing the non-shuffled (a) 1-channel and (b) 3-channels 2D collage representations of a 3D image volume. (c) An example 1-channel 2D collage image slice (512×512 pixel) containing 64 individual (non-shuffled) axial slices (64×64 pixel) of an actual kidney CT volume. The axially top and bottom slices (two corner slices in (c)) are colored to locate those in the randomly shuffled collages in (d)−(f).

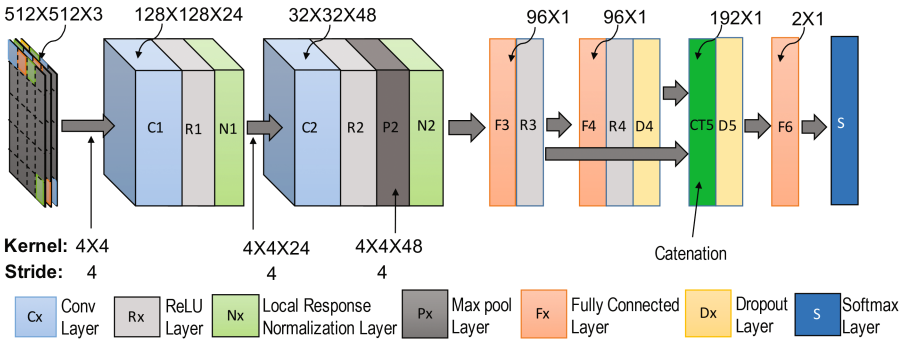

Fig. 2. The architecture of our collage deep convolutional neural network for pathological *vs.* healthy kidney classification. See Fig. 1 for the input image representation.

not applicable to all constituent axial slices. To address this challenge, we propose a novel approach where the slices within the 3D image are rearranged into an extended 2D image collage (Fig. 1). In a non-shuffled collage representation, each consecutive image slice (for 1-channel) or, a group of n consecutive image slices (for n-channels, where $n > 1$) along a particular direction are sequentially placed

on a 2D plane, which is schematically shown for a 1-channel and a 3-channels ($n = 3$) image in Fig. 1(a) and (b), respectively. Note that we opt to keep the collage dimension square (i.e. 512×512 pixel) in this experiment, however it is not a necessity. This collage not only ensures meaningful correspondence to the volume's single label but also allows for invaluable data augmentation by simple random reshuffling of image slices as well as by rotation and flipping. A non-shuffled 2D image collage representing an actual kidney CT data and its shuffle-based three augmented collages are shown in Fig. 1(c) and (d)−(f), respectively. Note that we prepare our CNN input data in a process shown in Fig. 1(b), where we set $n = 3$. The resulting dimension of a single CNN input data is $512 \times 512 \times 3$ pixel and the output was either 0 (healthy) or 1 (pathological).

2.3 Pathological *vs* Healthy Kidney Classification

CNN Architecture. Our proposed CNN has seven layers excluding the input. All of these layers except the 5th layer (concatenation layer) contained trainable weights.

Layer C1 is a convolutional layer that filters the input image with 24 kernels of size 4×4×3. Since we used collage-based image representation, we needed to carefully design our filter sizes and strides in a way that the convolutional (Cx) and max pooling (Px) filters do not overlap between two adjacent slices. To achieve this, we chose each edge size of the convolution filter to equal the stride in a particular layer. For example, the edge size of the convolution filter and the stride in the C1 layer were 4 and 4, respectively (Fig. 2). We chose a small convolutional filter size which tends to achieve better classification accuracy as demonstrated in [17].

Layer C2 is the second convolutional layer with forty eight 4×4×24 kernels applied to the output of C1. Unlike C1, we used a max pooling (P2) of 4×4×48 window in this layer to reduce the image size to 8×8 from 32×32.

The output of C2 is connected to a fully connected layer (F3), which contains 96 units. Similarly, a layer F4 contains 96 units and is fully connected to F3. We concatenated the units of F3 and F4 into CT5 in order to reduce possible information loss. This type of bypassing connections is typically suggested for better classification accuracy [18]. Note that the CT5 layer did not have any trainable weights.

The CT5 layer is connected to an F5 layer having 2 units. These units are connected to a softmax layer (S), which produces the relative probabilities for back-propagation and classification.

Solver. Our network was trained by minimizing the softmax loss between the desired and predicted labels. We used an optimization method called *Adam* [19]. All the parameters for this solver were set to the suggested default values, i.e. $\beta_1 = 0.9$, $\beta_2 = 0.999$, and $\epsilon = 10^{-8}$. We also employed a unit dropout (Dx) that drops 50% of units in both F4 and CT5 layers and used a weight decay of 0.005. The base learning rate was set to 0.01 and was decreased by a factor of

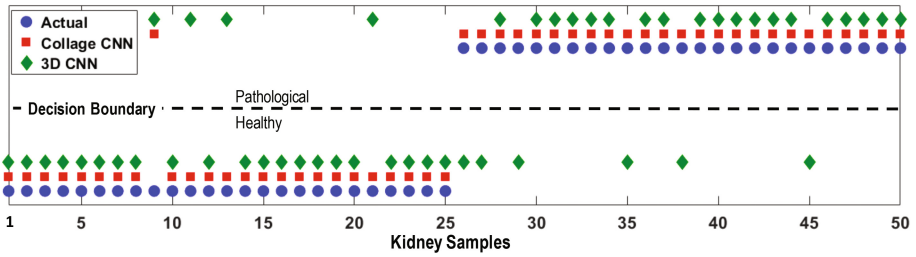

Fig. 3. Scatter plot showing the actual *vs.* predicted labels by the collage image-based and 3D CNNs.

0.1 to 0.0001 over 25000 iterations with a batch of 32 images processed at each iteration.

3 Results

We provide the classification accuracy results of our proposed collage image-based CNN as a bar plot in Fig. 3. We also compare our performance to that of a 3D CNN on the same plot. For the 3D CNN, we replaced the collage input (512×512×3 pixel) with the full 3D volume (64×64×64 pixel) of the kidney and performed 3D convolutions with a filter size of 4×4×64 with stride 4. For fair comparison, we chose the 3D volume dimension as 64×64×64 pixel, since each constituent axial slice in the collage was of 64×64 pixel. Other layer configurations remained the same as in Fig. 2. Both CNNs were implemented using *Caffe* [20]. The pre-processing of the data, visualizations, and comparisons were done in Matlab using the *MatCaffe* interface. Training was performed on a workstation with Intel 3.20 GHz Xeon processor, an Nvidia Quadro K600 GPU with 1 GB of VRAM, and 8 GB of host memory. Prior to generating the collage representation of the input data, we ensured a uniform voxel spacing in the image volume all axial, coronal and sagittal planes using interpolation. We manually defined the kidney ROI (sub)volume within the CT data in such a way that leaves an approximately 25% background area framing a kidney. Prior to training, both of the training and testing datasets were standardized.

In our experiments, we augmented the number of training samples by a factor of 40 by flipping and rotating the image slices as well as by random reshuffling the slice location within the collage. This augmentation process enabled by our novel image representation yielded a total of 4,400 2D image collages for training.

As demonstrated in Fig. 3, our proposed method succeeded in all but only one case out of 50 tested kidney samples, resulting in a classification accuracy of 98%. In comparison, the 3D convolution-based CNN failed in eight cases resulting in an accuracy of 80%.

Our preliminary results suggest that our proposed collage image representation may offer significant advantages for deep CNN-based classification tasks on 3D data. Our collage representation allows the convolution kernel to slide

over all the axial 2D slices in a 3D volume, which is impossible in case of a 3D CNN. The training time of the collage CNN was approximately 5 h (on our basic machine) while the 3D CNN took approximately 7 h to converge. We also augmented the 3D data by using data rotation and flipping before feeding to the 3D CNN, and the performance of the 3D CNN is expected to be better than our collage CNN. But because of the better augmentation capability of the collage representation, it performed better compared to the 3D CNN in our experiment. Thus, the collage representation seems best suited in the insufficient annotated medical data scenario. It is worth noting that in order to improve the classification accuracy by the 3D CNN approach, one may possibly have to increase the convolution kernel size and/or decrease the stride size in order to capture more features from the image volume. However, this would drastically increase the number of trainable weights, which would necessitate the use of expensive GPUs with large memory, and would cost more time to converge.

4 Conclusions

In this paper, we proposed a novel collage image representation within a CNN based classification scheme to enable deep learning from sparsely labelled 3D datasets. We applied our proposed method on CT abdominal scans from the TCIA database to discriminate healthy from cancerous kidneys containing renal cell carcinoma. Our method enables efficient 2D slice-based learning in the absence of slice-based labels. In addition, the proposed collage inherently allows for easy data augmentation through random reshuffling of the locations of image slices within the collage, thus facilitating more efficient training of the implicit relationship between bag labels and feature representation in weekly supervised ML settings. Our approach was shown to be impressively effective (98% classification accuracy) on weakly labeled data on a small sized data base of 160 kidney CTs outperforming 3D CNNs, though the latter's performance could potentially be improved with significant increase in labeled data as well as computation cost. In future work, we plan to couple an automatic kidney localization setup prior to proceed our proposed classification to produce a fully automated end-to-end kidney discrimination clinical tool.

Acknowledgement. This work is supported in part by the Institute for Computing, Information and Cognitive Systems (ICICS) at UBC.

References

1. Cancer Genome Atlas Research Network: Comprehensive molecular characterization of clear cell renal cell carcinoma. Nature **499**(7456), 43–49 (2013)
2. Siegel, R.L., Miller, K.D., Jemal, A.: Cancer statistics. Cancer J. Clin. **65**(1), 5–29 (2015)
3. Ridge, C.A., Pua, B.B., Madoff, D.C.: Epidemiology and staging of renal cell carcinoma. Semin. Interv. Radiol. **31**(01), 003–008 (2014)

4. Escudier, B., Eisen, T., Porta, C., Patard, J.J., Khoo, V., Algaba, F., Mulders, P., Kataja, V., ESMO Guidelines Working Group: ESMO Clinical Practice Guidelines for diagnosis, treatment and follow-up. Ann. Oncol. 23(suppl 7), vii65-vii71 (2012)
5. Criminisi, A., Shotton, J., Robertson, D., Konukoglu, E.: Regression forests for efficient anatomy detection and localization in CT studies. In: Menze, B., Langs, G., Tu, Z., Criminisi, A. (eds.) MCV 2010. LNCS, vol. 6533, pp. 106–117. Springer, Heidelberg (2011). doi:10.1007/978-3-642-18421-5_11
6. Criminisi, A., Robertson, D., Konukoglu, E., Shotton, J., Pathak, S., White, S., Siddiqui, K.: Regression forests for efficient anatomy detection and localization in computed tomography scans. Med. Image Anal. **17**(8), 1293–1303 (2013)
7. Cuingnet, R., Prevost, R., Lesage, D., Cohen, L.D., Mory, B., Ardon, R.: Automatic detection and segmentation of kidneys in 3D CT images using random forests. In: Ayache, N., Delingette, H., Golland, P., Mori, K. (eds.) MICCAI 2012. LNCS, vol. 7512, pp. 66–74. Springer, Heidelberg (2012). doi:10.1007/978-3-642-33454-2_9
8. Lu, X., Xu, D., Liu, D.: Robust 3D organ localization with dual learning architectures and fusion. In: Carneiro, G., Mateus, D., Peter, L., Bradley, A., Tavares, J.M.R.S., Belagiannis, V., Papa, J.P., Nascimento, J.C., Loog, M., Lu, Z., Cardoso, J.S., Cornebise, J. (eds.) LABELS/DLMIA -2016. LNCS, vol. 10008, pp. 12–20. Springer, Cham (2016). doi:10.1007/978-3-319-46976-8_2
9. Hussain, M.A., Hamarneh, G., O'Connell, T.W., Mohammed, M.F., Abugharbieh, R.: Segmentation-free estimation of kidney volumes in CT with dual regression forests. In: Wang, L., Adeli, E., Wang, Q., Shi, Y., Suk, H.-I. (eds.) MLMI 2016. LNCS, vol. 10019, pp. 156–163. Springer, Cham (2016). doi:10.1007/978-3-319-47157-0_19
10. Dietterich, T.G., Lathrop, R.H., Lozano-Pérez, T.: Solving the multiple instance problem with axis-parallel rectangles. Artif. Intell. **89**(1), 31–71 (1997)
11. Zhu, W., Lou, Q., Vang, Y.S., Xie, X.: Deep Multi-instance Networks with Sparse Label Assignment for Whole Mammogram Classification. arXiv preprint arXiv:1612.05968 (2016)
12. Yan, Z., Zhan, Y., Peng, Z., Liao, S., Shinagawa, Y., Zhang, S., Metaxas, D.N., Zhou, X.S.: Multi-instance deep learning: discover discriminative local anatomies for bodypart recognition. IEEE Trans. Med. Imaging **35**(5), 1332–1343 (2016)
13. Xu, Y., Mo, T., Feng, Q., Zhong, P., Lai, M., Eric, I., Chang, C.: Deep learning of feature representation with multiple instance learning for medical image analysis. In: IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), pp. 1626–1630 (2014)
14. Kraus, O.Z., Ba, J.L., Frey, B.J.: Classifying and segmenting microscopy images with deep multiple instance learning. Bioinformatics **32**(12), i52–i59 (2016)
15. Clark, K., Vendt, B., Smith, K., Freymann, J., Kirby, J., Koppel, P., Moore, S., Phillips, S., Maffitt, D., Pringle, M., Tarbox, L., Prior, F.: The cancer imaging archive (TCIA): maintaining and operating a public information repository. J. Digit. Imaging **26**(6), 1045–1057 (2013)
16. Shinagare, A.B., Vikram, R., Jaffe, C., Akin, O., Kirby, J., Huang, E., Freymann, J., Sainani, N.I., Sadow, C.A., Bathala, T.K., Rubin, D.L.: Radiogenomics of clear cell renal cell carcinoma: preliminary findings of the cancer genome atlas–renal cell carcinoma (TCGA-RCC) imaging research group. Abdom. Imaging **40**(6), 1684–1692 (2015)
17. Sermanet, P., Eigen, D., Zhang, X., Mathieu, M., Fergus, R., LeCun, Y.: OverFeat: integrated recognition, localization and detection using convolutional networks. In: Proceedings of ICLR (2014)

18. Sun, Y., Wang, X., Tang, X.: Deep learning face representation from predicting 10,000 classes. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 1891–1898 (2014)
19. Kingma, D., Ba, J.: Adam: a method for stochastic optimization. In: 3rd International Conference for Learning Representations (2015)
20. Jia, Y., et al.: Caffe: convolutional architecture for fast feature embedding. In: ACM International Conference on Multimedia, pp. 675–678 (2014)