IEEE ICIP 2019

"A dual attention dilated residual network for liver lesion classification and localization on CT images "

Xiao Chen, Lanfen Lin, Dong Liang, Hongjie Hu, Qiaowei Zhang, Yutaro Iwamoto,

Xian-Hua Han, Yen-Wei Chen, Ruofeng Tong, Jian Wu.

Zhejiang University, Ritsumeikan University, Sir Run Run Shaw Hospital

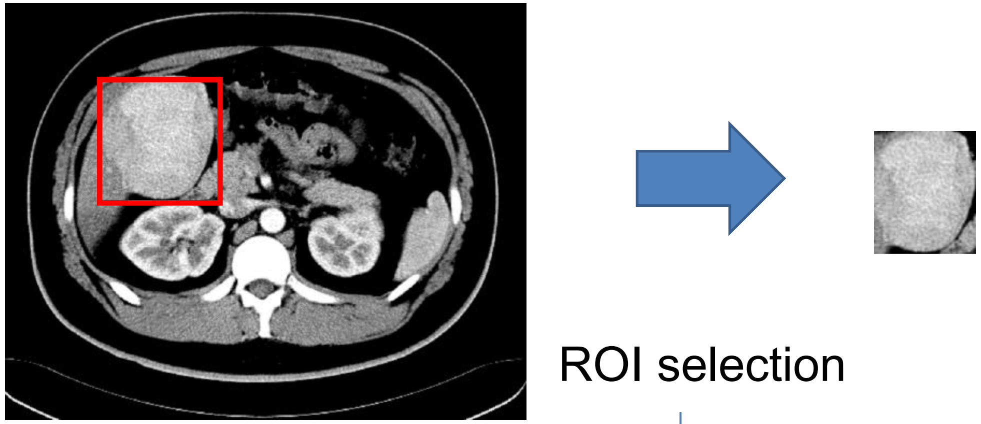
Content

Introduction

Background

Liver cancer is the second most common cause of cancer-related deaths among men and sixty among women.

Major concern limits automatic liver lesion classification is that previous methods are conducted on lesion level, which relies heavily on ROI selection process.



ROI selection

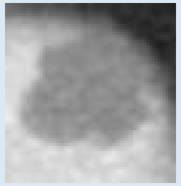
Labor-intensive
Manual annotations

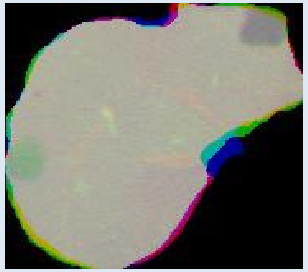
Automatic lesion
detection/segmentation

Introduction

Motivation

ART Phase





Previous methods:

1. segment liver lesions
2. conduct lesion-level classification
   (ROI-level, patch-level, or both)

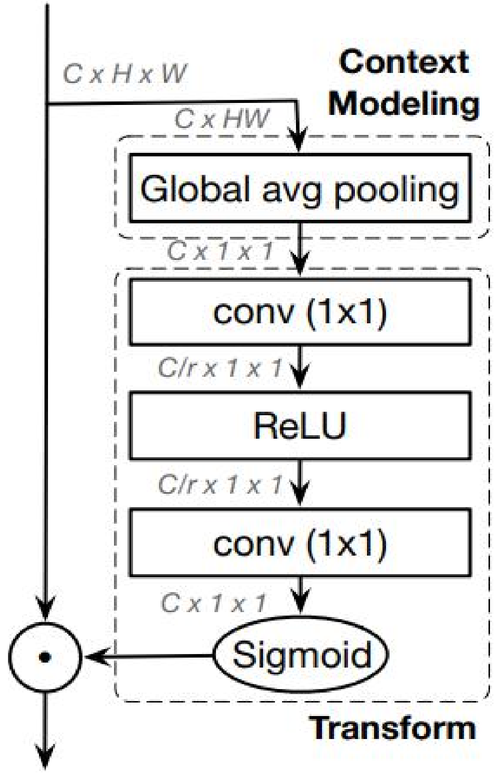
Our proposed methods:

1. segment whole liver area
2. conduct image-level classification
   (without lesion detection or segmentation)

To relieve the burden of expensive pixel-level lesions' annotations, we first explored the potential of using the whole liver slice image for liver lesion classification without pre- detection or pre-selection of the ROI.

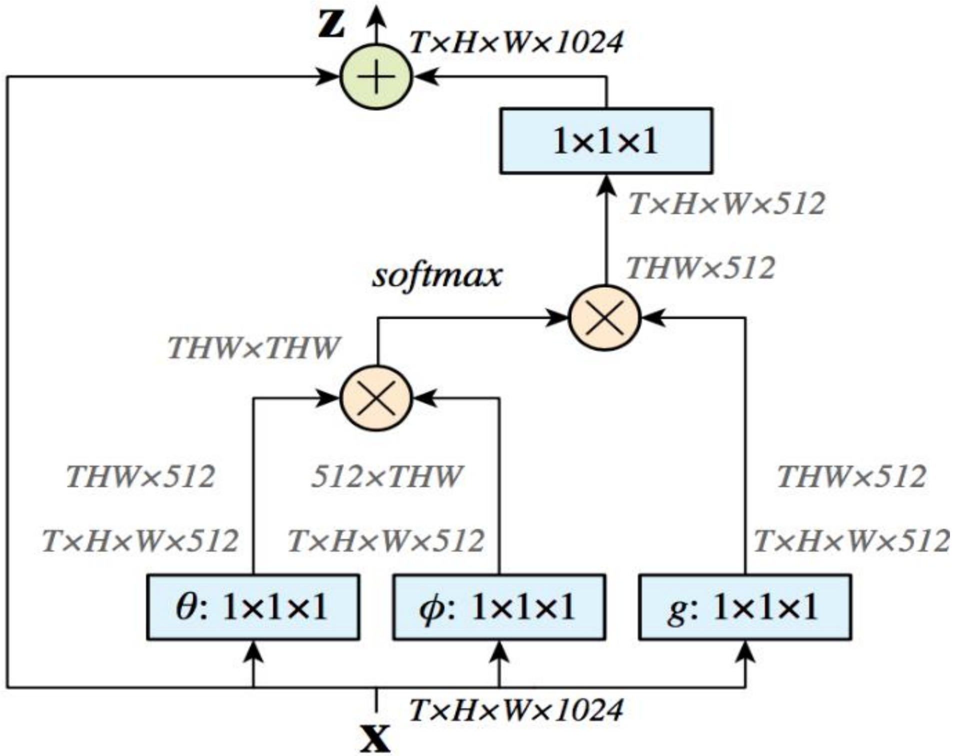
Related work

➢ Attention mechanism in Computer Vision



Squeeze-Excitation block[1]

(1) explicitly model channel-interdependencies



Non-local block[2]

(2) model spatial-temporal dependencies

[1] Hu, Jie, Li Shen, and Gang Sun. "Squeeze-and-excitation networks." Proceedings of the IEEE conference on computer vision and pattern recognition. 2018.
[2] Wang, Xiaolong, et al. "Non-local neural networks." Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2018.

Introduction

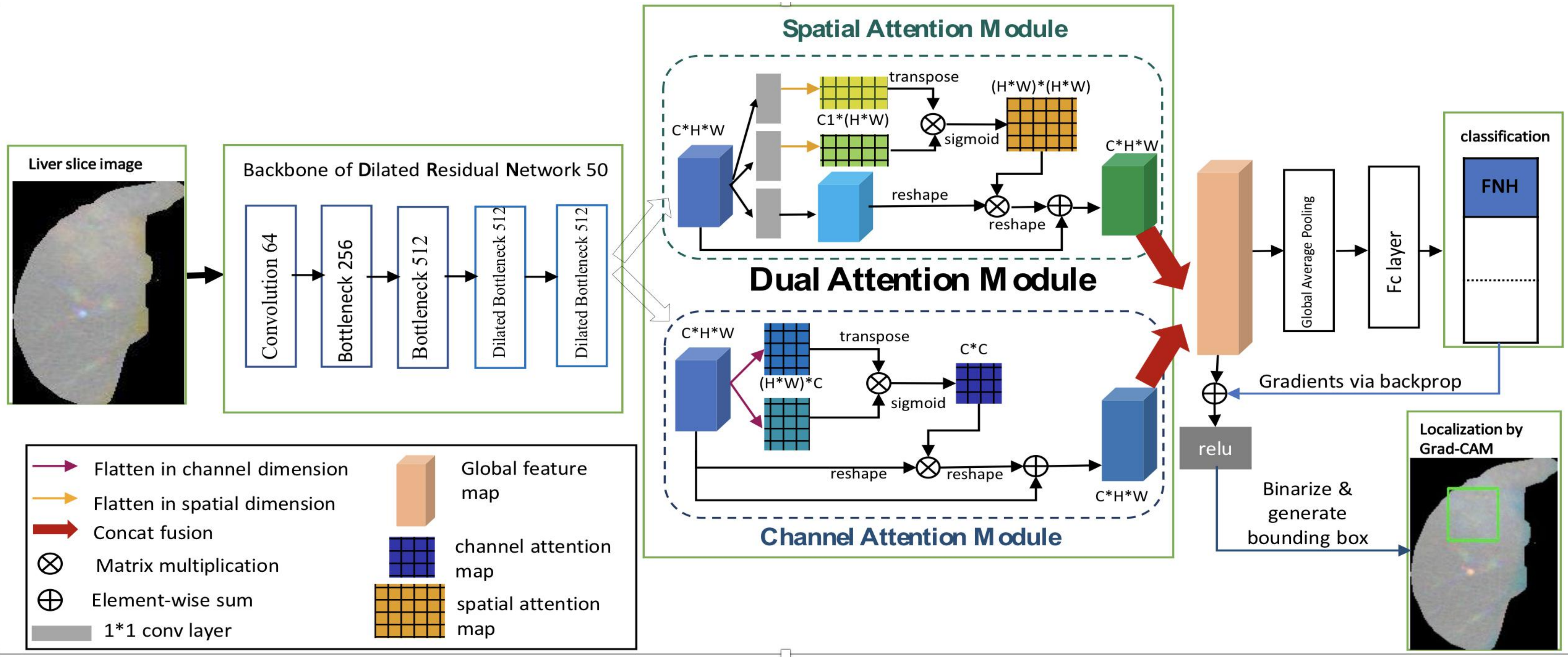
Contributions

- We proposed the DADRN framework which no longer relies on lesion annotations and could tackle the lesion classification problem as a one-stage process.
- Our dual-attention mechanism integrates similar features of high-level feature map from a global view, which improves DRN’s lesion recognition performance
- The experimental results show that DADRN is comparable to the ROI-level classification model and is superior to other state-of-the-art attention-based classification models in lesion classification task and weakly- supervised lesion localization task.

Content

Methodology

Overview of framework

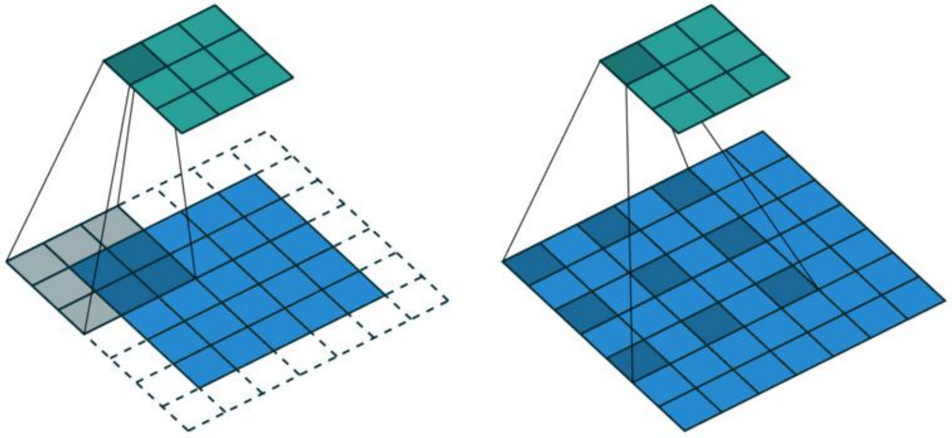


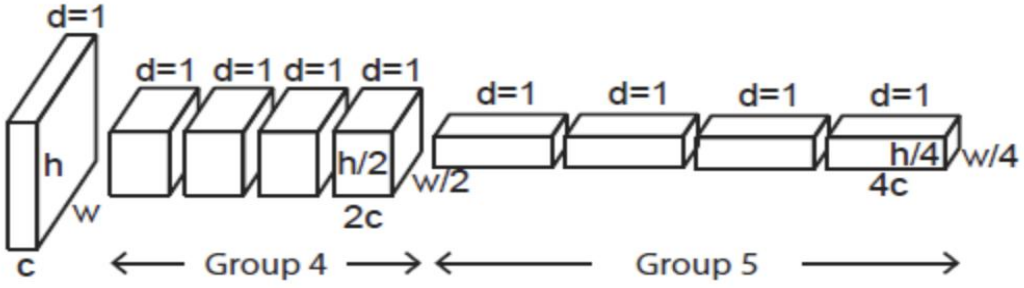
Methodology

Backbone Network

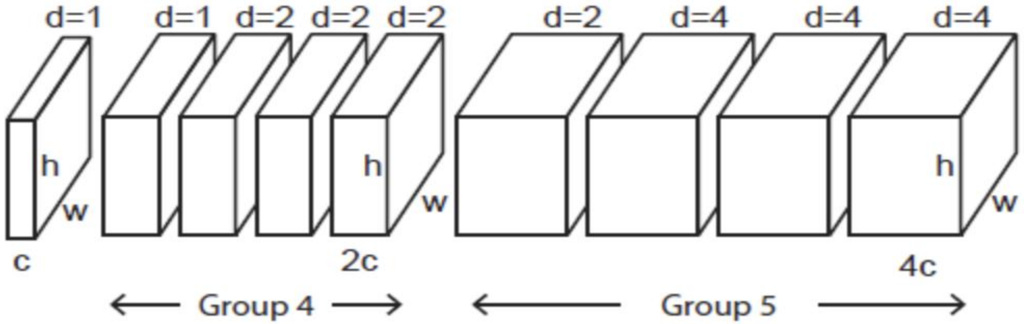
- Dilated Residual Network (DRN) (Yu et al. 2017)

DRN is chosen as the backbone classification network. Since the output of Group5 in DRN is 28*28, which is much larger than that of original Resnet.





(a) ResNet

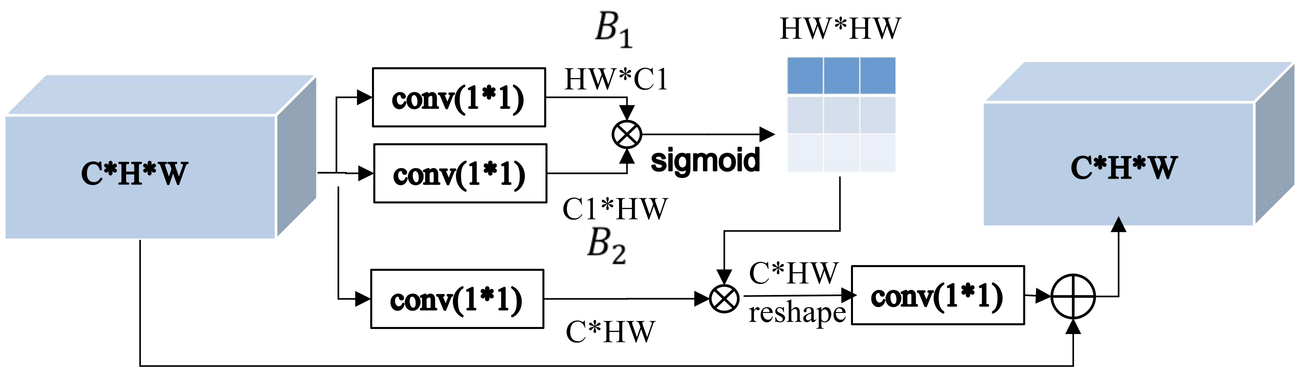


(b) DRN

Methodology

Our Dual attention block

- Closer look to dual attention block



a. spatial attention block [1]

Step1: Batch matrix multiplication

$$S_{ij} = {B_1}^T B_2$$

Step2: Normalize similarity map

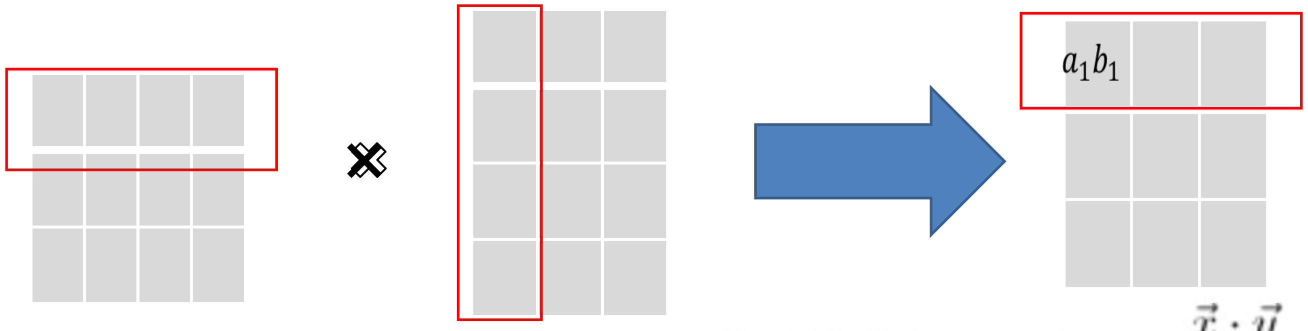
$$\alpha_{ji} = \frac{1}{1+e^{-S_{ij}}},$$

Step3: Synthesize new feature map

$$o_j = \sum_{i=1}^{N} \alpha_{j,i} D_i$$

Step4: Adaptively learn the weight of synthesized feature map

$$y_j = \beta o_j + A_j$$



$$sim(X, Y) = cos\theta = \frac{\vec{x} \cdot \vec{y}}{||x|| \cdot ||y||}$$

b. Illustration of step1 batch matrix multiplication

[1] Wang, Xiaolong, et al. "Non-local neural networks." Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2018.

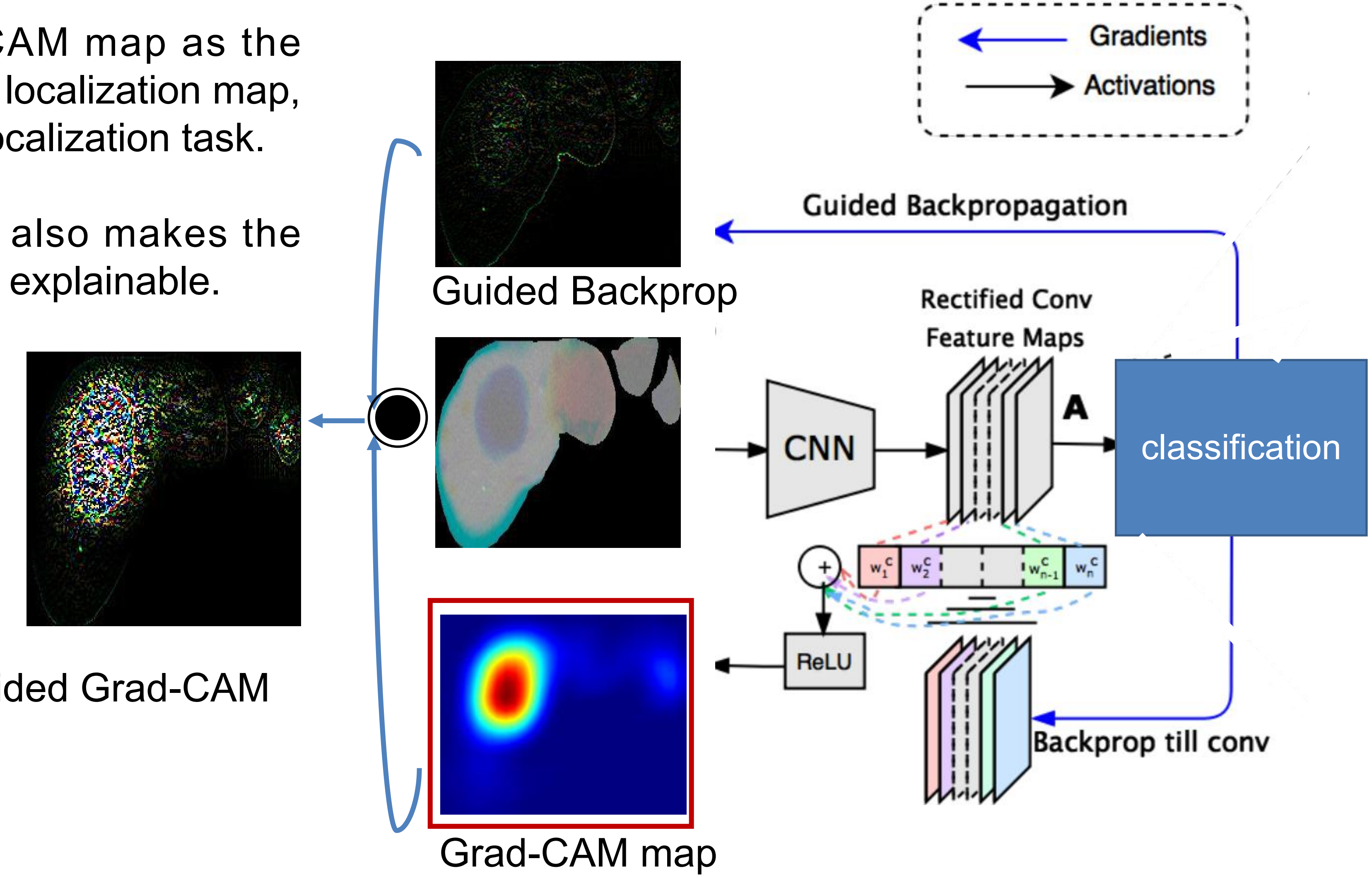
Methodology

Visualization of Attention Maps

➢ Gradient-weighted Class Activation Maps (Grad-CAM) (Selvaraju et al. 2017)

- Here, we adopt Grad-CAM map as the weakly-supervised lesion localization map, which is used for lesion localization task.
- The visualization result also makes the classification model more explainable.

$$L^{s}_{GradCAM} = ReLU\left(\sum_t \overbrace{\frac{1}{z}\sum_i\sum_j \frac{\partial y^s}{\partial A^k_{ij}}}^{GlobalAveragePooling} A^t\right)$$



Content

Experiments

Dataset

A total of **1091 CT liver slice images** in the arterial phase.
Five types: normal, CYST, FNH, HCC and HEM.
To leverage 3D context information, each liver slice image contains two pieces of neighboring. The input images were all resized to $\mathbf{224} \times \mathbf{224} \times \mathbf{3}$. To eliminate the effect of randomness, we split our dataset twice and the patient case did not overlap among the train set, validation set and test set.

Type	Train		Validation		Test		Total
	Set1	Set2	Set1	Set2	Set1	Set2	
Normal	135	126	41	57	51	44	227
CYST	168	166	56	59	69	68	293
FNH	75	75	29	27	26	28	130
HCC	149	143	52	57	50	51	251
HEM	112	114	38	37	40	39	190

Experiments

Comparison to other methods

Compared with other attention-based CNN, baseline DRN, state-of-the-art ROI-level lesion classification method (ResGLNet). Our DADRN50-B is superior in most cases and closed to ROI-level method.

1. Comparison of class-wise classification accuracy

	Method	Normal	CYST	FNH	HCC	HEM
	DRN50 [18]	0.9788	0.9327	0.7596	0.8427	0.5278
	SEResnet50[14]	0.9334	0.9327	0.7788	0.9116	0.5917
	RAResnet50[13]	0.9675	0.9182	0.7596	0.8227	0.5556
1	**SADRN50-A**	0.9577	0.9096	0.8132	0.8816	0.6625
	SADRN50-B	0.9334	0.8761	0.7775	0.8220	0.5458
1	**CADRN50-A**	0.9675	0.9551	**0.8530**	0.9016	0.6181
	CADRN50-B	0.9588	0.9413	0.8324	0.8322	0.5847
2	**DADRN50-A**	0.9690	0.9451	0.7802	0.8024	0.7069
	DADRN50-B	**0.9804**	0.9551	0.8159	**0.9116**	0.6819
	ResGLNet [21]	-	**0.9615**	0.8405	0.8846	**0.8462**

① Different normalization strategy in dual attention block: sigmoid(A) softmax(B)

② Different fusion strategy of spatial and channel attention: sum fusion(A) concatenate fusion(B)

Experiments

Comparison to other methods

Compared with Image-level methods, Our DADRN50-B is superior in all 5-class classification metrics (include normal liver slice images).

Comparison of 5-class overall classification performance

Method	Accuracy	F1	Precision	Recall
DRN50 [18]	0.8083	0.8197	0.8294	0.8207
SEResnet50 [14]	0.8296	0.8265	0.8552	0.8149
RAResnet50 [13]	0.8047	0.8041	0.8304	0.7905
SADRN50-A	0.8449	0.8372	0.8463	0.8346
CADRN50-A	0.8591	0.8263	0.8506	0.8149
DADRN50-A	0.8407	0.8213	0.8446	0.8111
DADRN50-B	**0.8690**	**0.8412**	**0.8528**	**0.8386**

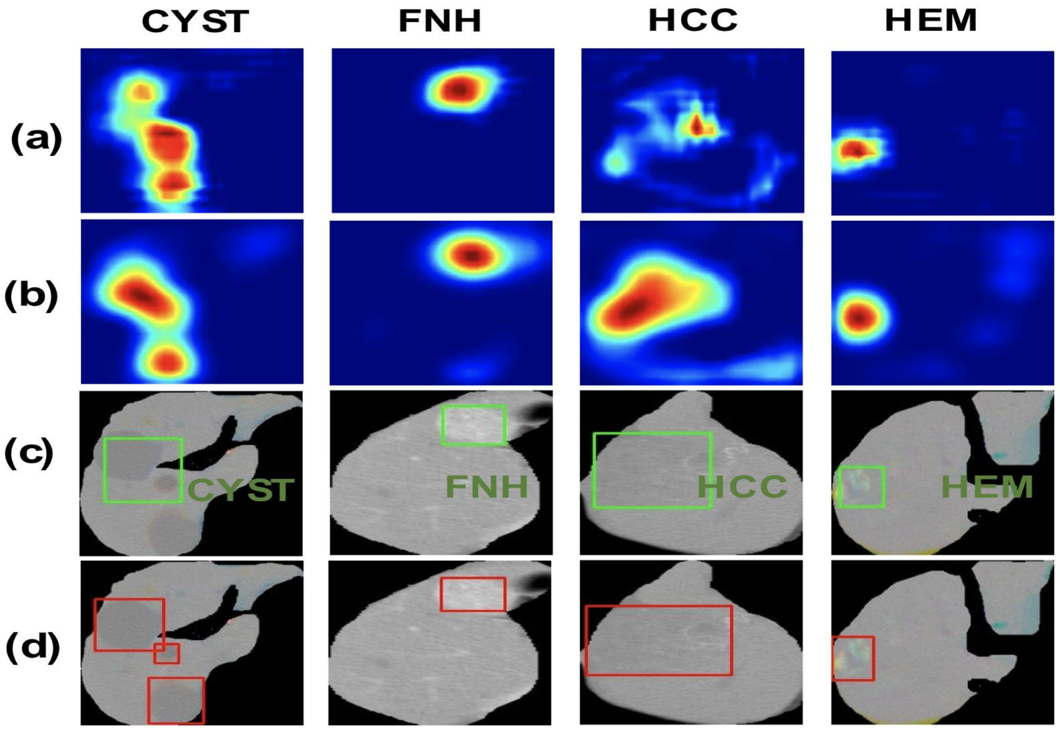
Experiments

Weakly-supervised localization performance

Compared with the state-of-the-art attention-based CNN and baseline DRN, our DADRN50-B is much better in lesion localization task.

$$Localization\ Acc_c = \frac{total\ num\ of\ correct\ localized\ slices\ for\ class\ c}{total\ num\ of\ slices\ for\ class\ c}$$

Method	CYST	FNH	HCC	HEM
DRN50 [18]	0.5110	0.6676	0.5941	0.3798
SEResnet50 [14]	0.1898	0.0742	0.7327	0.2532
RAResnet50 [13]	0.2628	0.0742	0.6931	0.3292
DADRN50-B	**0.5986**	**0.6676**	**0.7327**	**0.5064**



(a) Grad-CAM map of DRN; (b) Grad-CAM map of DADRN; (c) weakly-supervised localization result generated by (b); (d) ground truth of each slice image.

Content

Conclusion

- Our proposed method allows for implementing lesion classification without pre-detection or pre-selection of lesion ROIs.
- Dual attention module improve DRN’s lesion recognition ability
- DADRN is comparable to state-of-the-art ROI-level classification method and is superior to most state-of-the-art attention-based methods in lesion classification task and weakly-supervised lesion localization task.

Thank you