Integrating Medical Knowledge into Deep Learning Architectures

Multimedia Lab, Computer Science and Engineering, IIT Guwahati

Akshay Daydar
Research Scholar, IITG

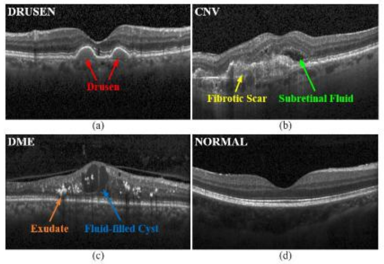
The Need of Deep Learning in Medical Imaging

Practical perspective: To help the clinicians for dependable medical assistance.

Research perspective: Medical images involves visual and latent information of disease biomarkers that can be traced using deep learning.

Tasks : Disease diagnosis, Lesion, organ, and abnormality detection, Lesion, organ segmentation, medical image registration, medical image retrieval, Medical report generation.

Challenges: Medical images are grayscale in nature, have non-differential spatial context and often contains small ROI compared to image dimensions.



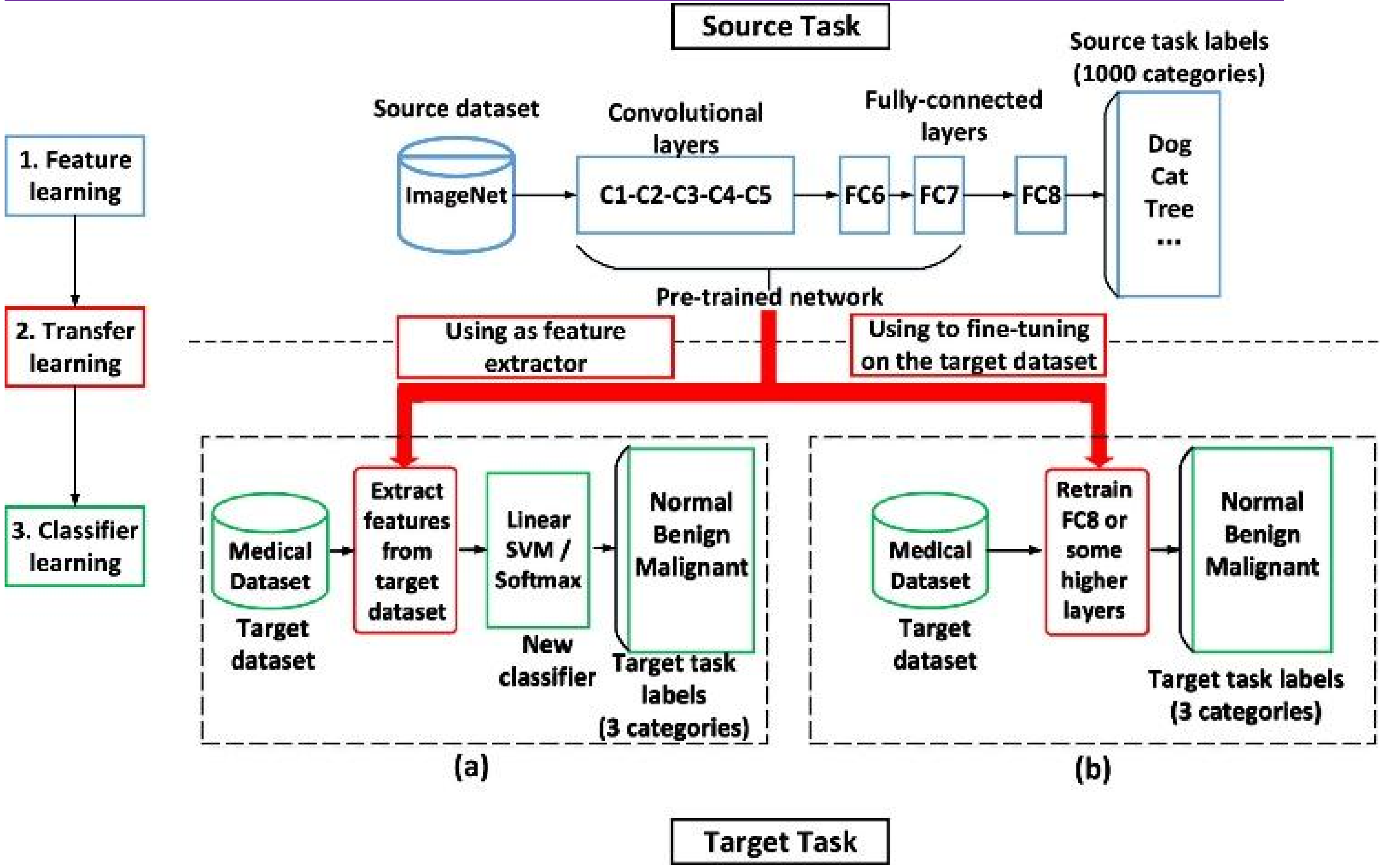
OCT image


MRI

Xie, Xiaozheng, et al. "A survey on incorporating domain knowledge into deep learning for medical image analysis." *Medical Image Analysis* 69 (2021): 101985.

Recent Trends

Training Pattern for Medical Images



Two strategies to utilize the pre-trained network on natural images: (a) as a feature extractor and (b) as an initialization which will be fine-tuned on the target dataset [Xie et al.].

1. The training pattern
2. The general diagnostic patterns they view images- MRI, CT
3. The areas on which they usually focus- knee xray
4. The features (e.g characteristics, structures, shapes) they give special attention to, and - Cancer
5. Other related information for diagnosis

Xie, Xiaozheng, et al. "A survey on incorporating domain knowledge into deep learning for medical image analysis." Medical Image Analysis 69 (2021): 101985.

Recent Trends

Incorporating domain knowledge to Classification

1. Training Deep learning model with disease specific descriptors

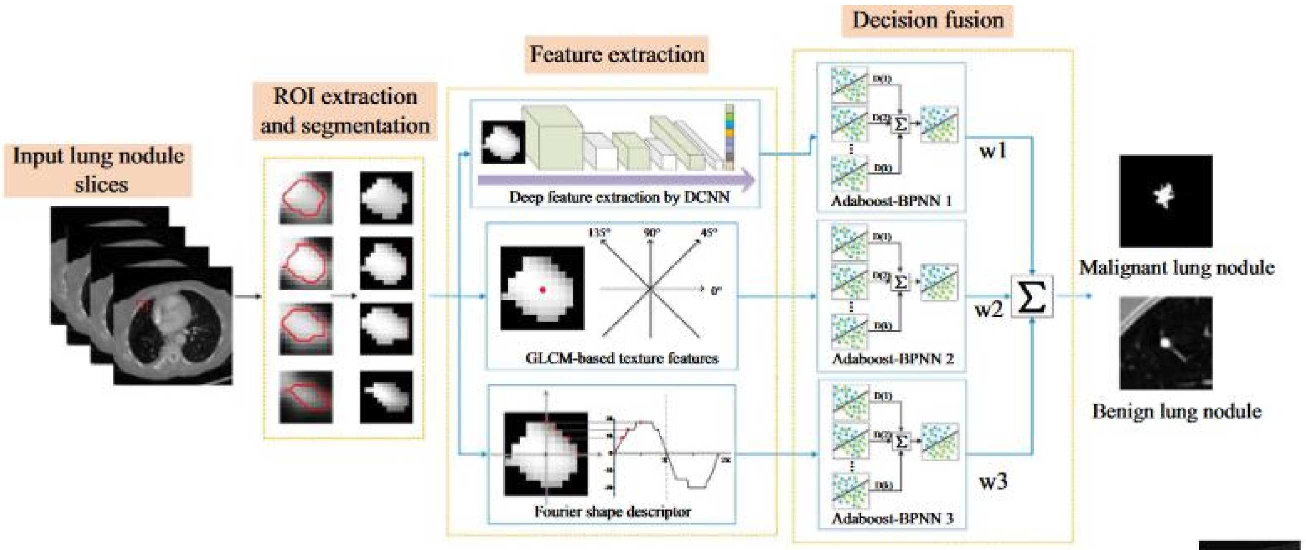
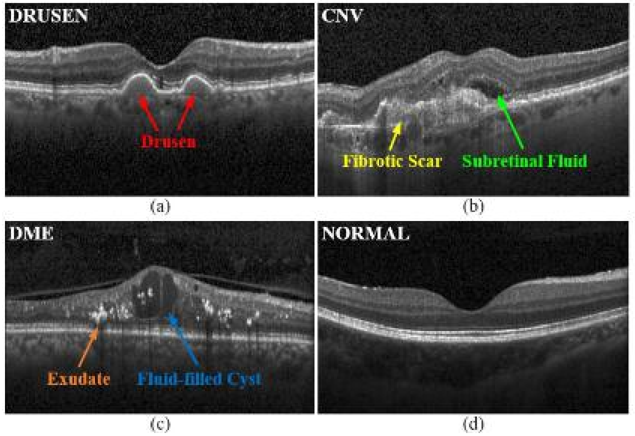
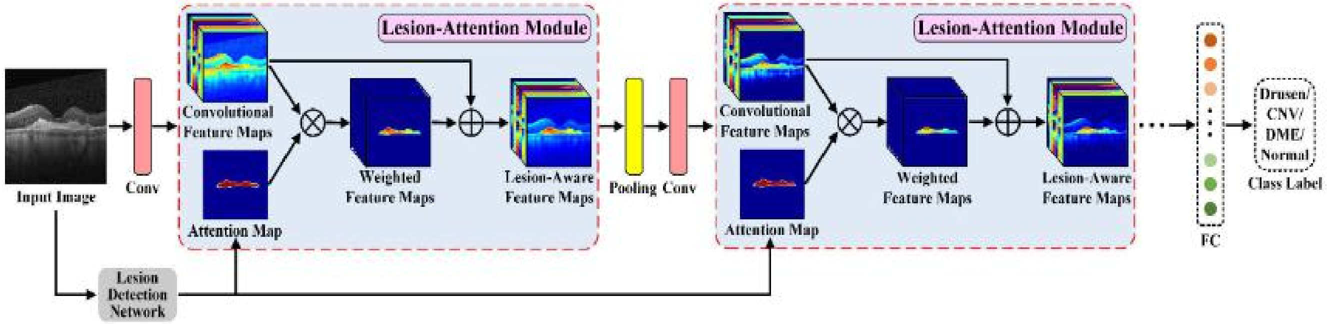


Diagram of Fuse-TSD lung nodule classification algorithm.

Texture and shape descriptors such as gray level co-occurance matrix, faret shape measure, moment invariants, point distance histograms and fourier descriptors can be used to characterize heterogeneous shape of the nodules [Xie et al.]

2. Utilizing Attention Mechanism for guiding the classification tasks





The framework of LACNN for retinal OCT image classification [Fang et al.]

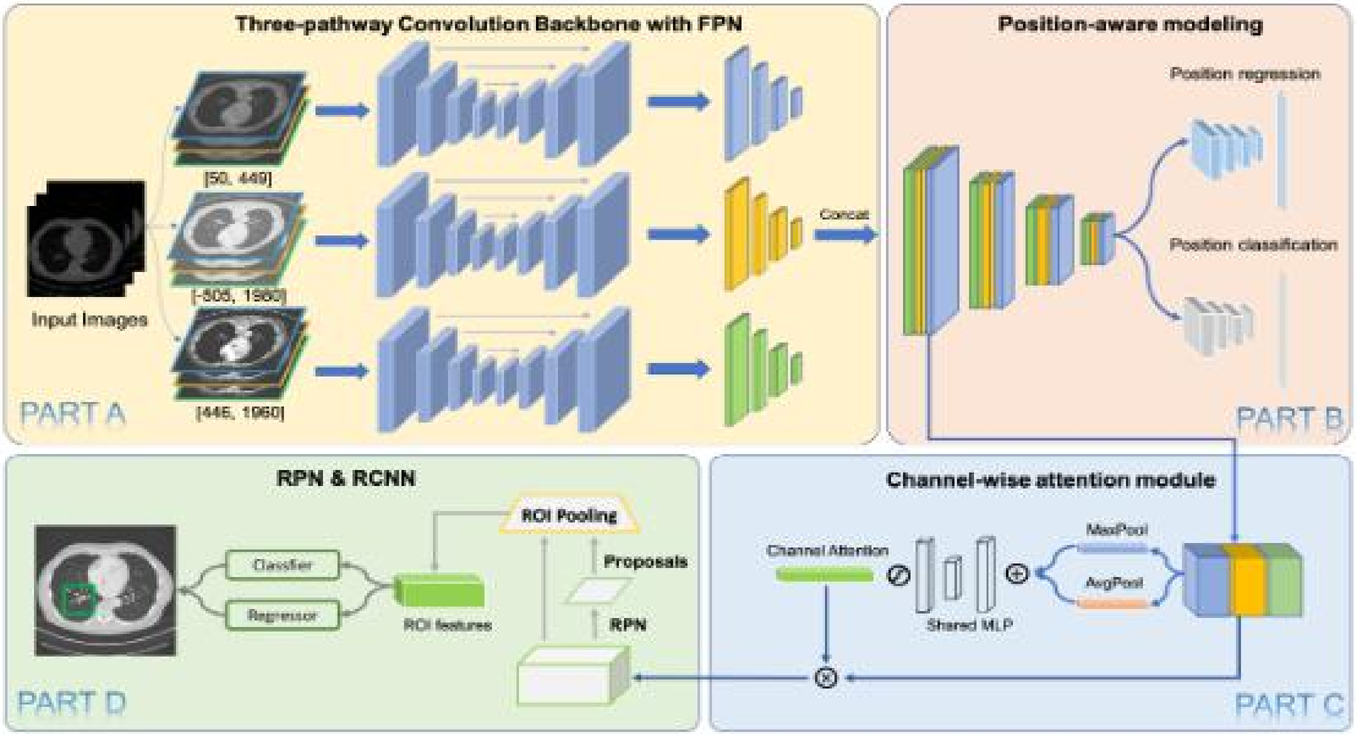
Xie, Yutong, et al. "Fusing texture, shape and deep model-learned information at decision level for automated classification of lung nodules on chest CT." Information Fusion 42 (2018): 102-110.

Fang, Leyuan, et al. "Attention to lesion: Lesion-aware convolutional neural network for retinal optical coherence tomography image classification." IEEE transactions on medical imaging 38.8 (2019): 1959-1970.

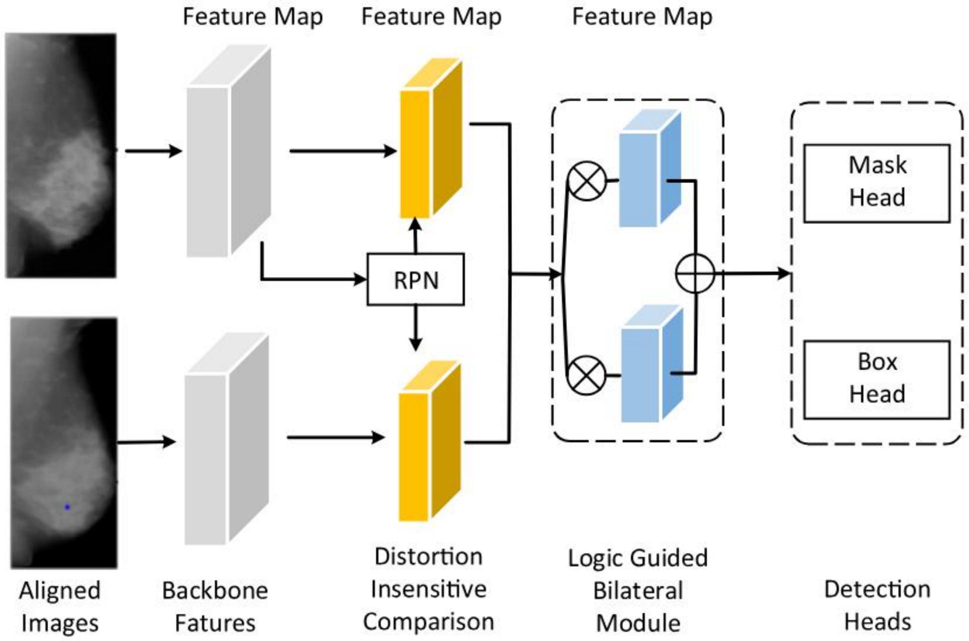
Recent Trends

Incorporating domain knowledge to Object Detection

3. Utilizing images under different settings/Multiple views/modalities/analyzing adjacent slides for computer vision tasks



Overview of MVP-Net. Coarser feature maps of FPN are omitted in part C and D for clarity, they use the same attention module with shared parameters for feature aggregation [Li et al.]



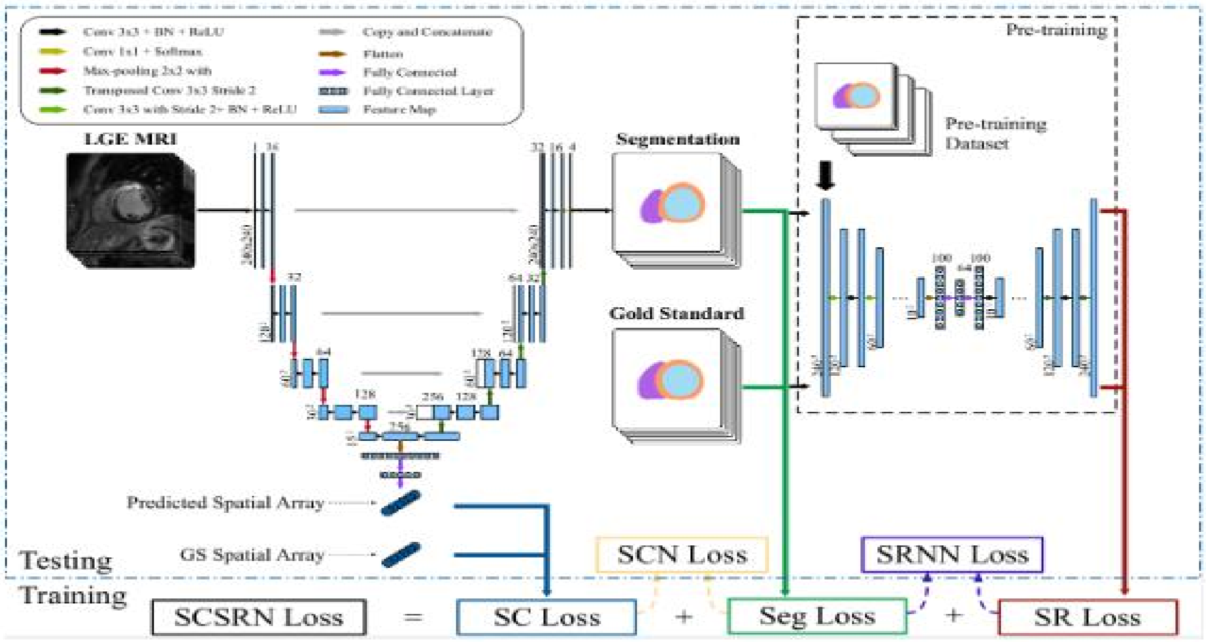
The workflow of mammogram mass detection by integrating the bilateral information (Liu et al., 2019), where the aligned images are fed into two networks seperately to extract features for further detection [Liu et al.]

Li, Zihao, et al. "MVP-Net: multi-view FPN with position-aware attention for deep universal lesion detection." International Conference on Medical Image Computing and Computer-Assisted Intervention. Springer, Cham, 2019.
Liu, Yuhang, et al. "From unilateral to bilateral learning: Detecting mammogram masses with contrasted bilateral network." *Medical Image Computing and Computer Assisted Intervention–MICCAI 2019: 22nd International Conference, Shenzhen, China, October 13–17, 2019, Proceedings, Part VI 22*. Springer International Publishing, 2019.

Recent Trends

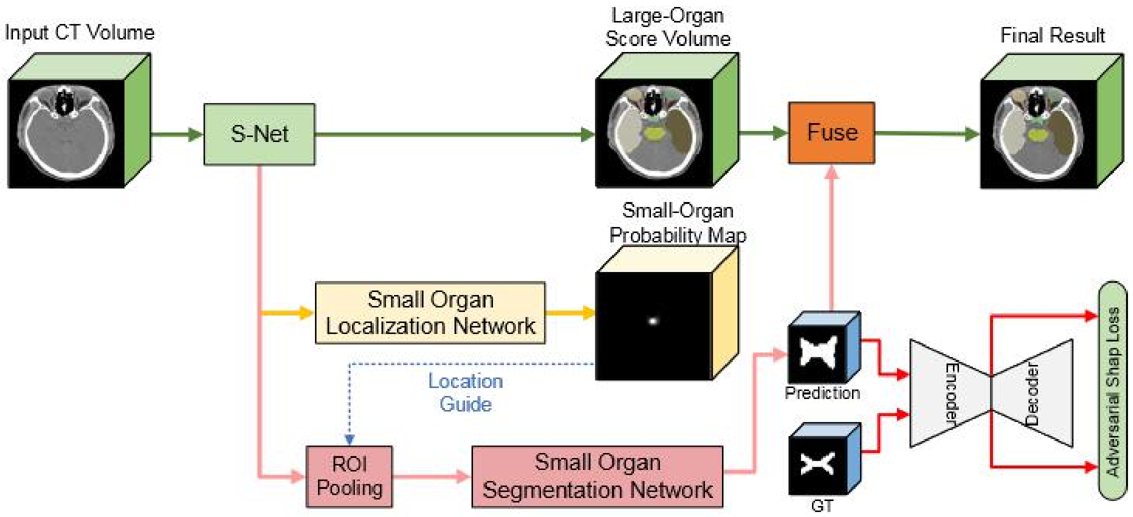
Incorporating domain knowledge to Segmentation

4. Utilizing loss function as regularizing term for fine tuning the Segmentation masks



Overall structure of SRSCN, whose loss comes from three parts: the segmentation loss is specially design as a function of cross entropy and Dice, the spatial constraint (SC) loss to assist segmentation, and the shape reconstruction (SR) loss for shape regularization [Yue et al.]

5. Utilizing localization network to locate small tissues for fine tuning the segmentation results

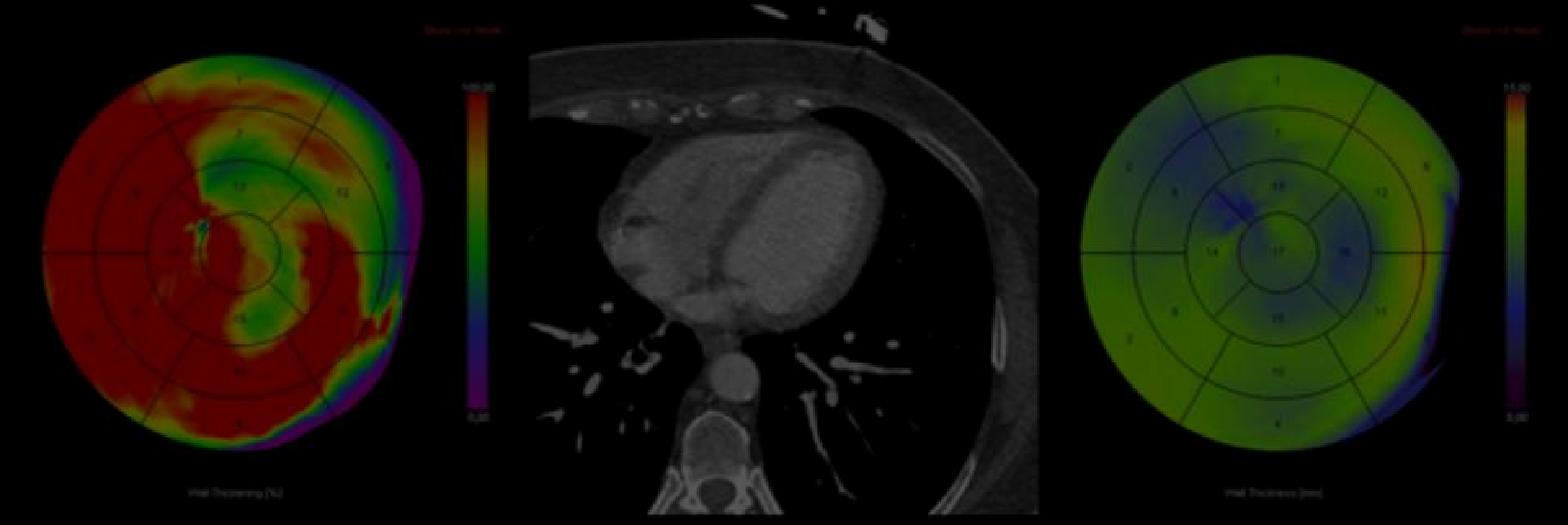


Overall framework - FocusNetv2 [Gao et al.]

Yue, Qian, et al. "Cardiac segmentation from LGE MRI using deep neural network incorporating shape and spatial priors." *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer, Cham, 2019.
Gao, Yunhe, et al. "FocusNetv2: Imbalanced large and small organ segmentation with adversarial shape constraint for head and neck CT images." *Medical Image Analysis* 67 (2021): 101831..

Key Concepts in Medical Imaging

- Multiple Instance Learning
- Medical Report Generation



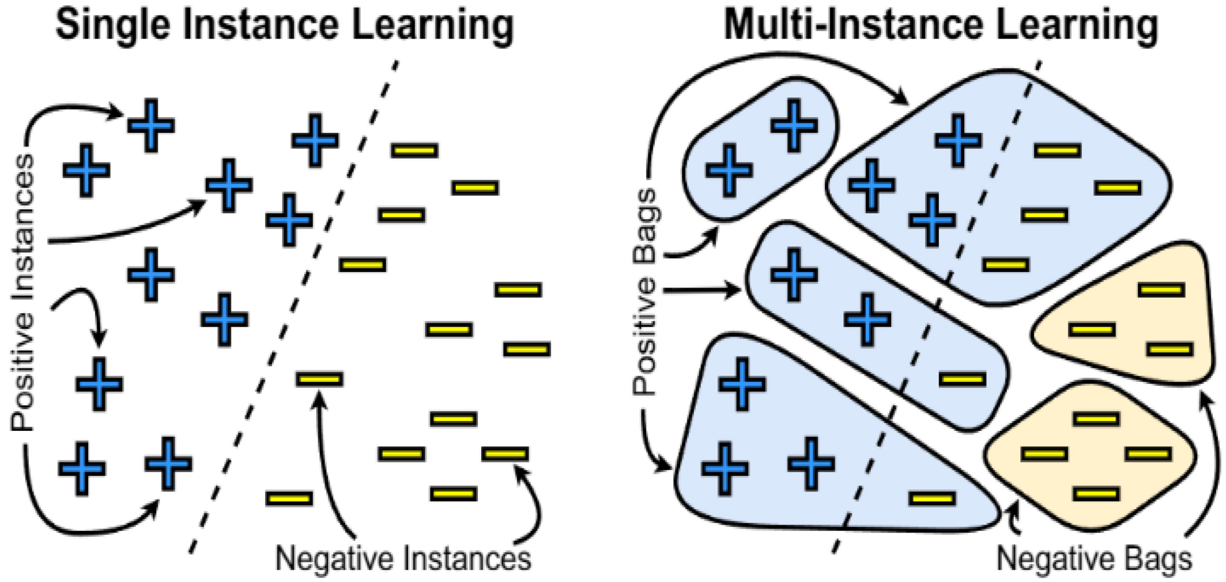
Multiple Instance Learning

In MIL, each training sample is a **bag** (e.g., a whole medical image), and it contains **multiple instances** (e.g., image patches or regions). Labels are **only provided at the bag level**, and the task is to learn from these to make both **bag-level and potentially instance-level** predictions.

- **Positive bag**: At least one instance is positive.
- **Negative bag**: All instances are negative.

Histopathology (e.g., cancer detection in

CT/MRI Volumes



https://www.mdpi.com/2079-9292/12/20/4323

Multiple Instance Learning

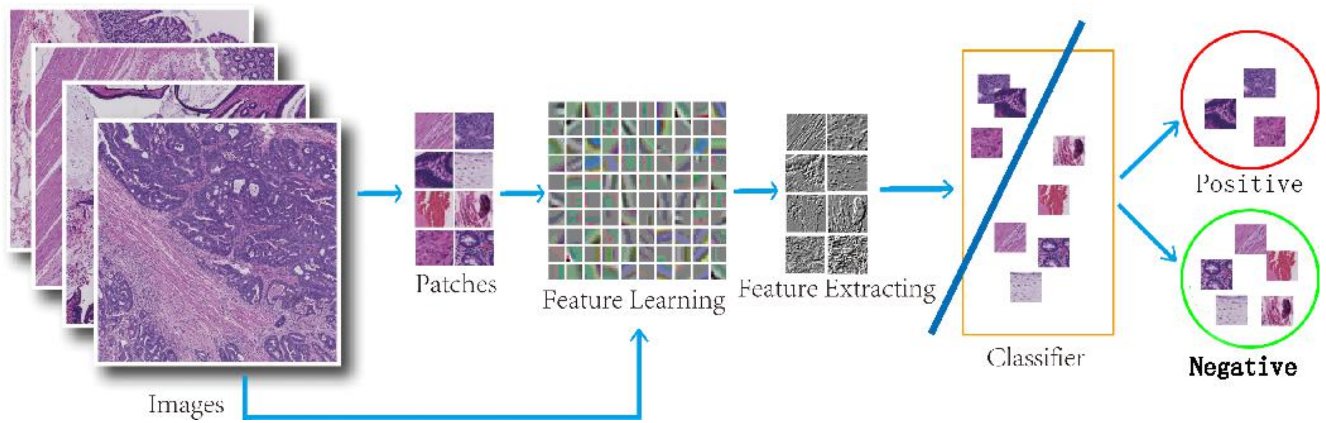
Bag of instances $X=\{\mathbf{x}_1,\cdots,\mathbf{x}_K\}$ with a bag label y, where $\mathbf{x}_k \in \mathbb{R}_D$ is the k-th instance.

The number of instances per bag K may vary across samples.

In its standard formulation, the instances of a bag exhibit neither dependency nor ordering among each other.

It is further assumed that binary instance labels $y_k \in \{0,1\}$ exist but are not necessarily known.

The binary bag label is 1 if and only if at least one instance label is 1, i.e., $y = \max_k\{y_k\}$.



https://arxiv.org/html/2406.04280v1

Medical Report Generation

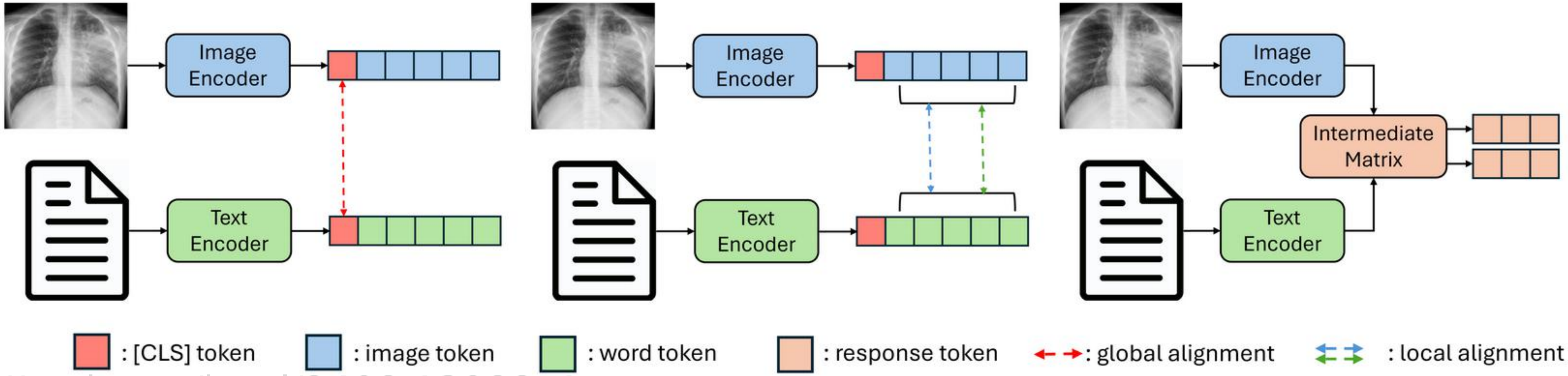
Medical Report Generation refers to the automatic creation of clinical or diagnostic reports from medical data using artificial intelligence (AI) or machine learning (ML) techniques.

In detail:

• **Input**: Medical images (e.g., X-rays, MRIs), patient records, or sensor data.

• **Output**: A written report that describes the findings, diagnosis, or recommendations (e.g., "No signs of pneumonia" or "Fracture detected in the left femur").

Impression: no acute radiographic cardiopulmonary process.
Findings: Three images are available for review. the heart size is normal. the mediastinal contour is within normal limits. the lungs are free of any focal infiltrates. there are no nodules or masses. no visible pneumothorax. no visible pleural fluid. the xxxx are grossly normal. there is no visible free intraperitoneal air under the diaphragm.



https://arxiv.org/html/2408.13988v1

Figure: Flowcharts of three representative alignment methods.

Medical Report Generation

Given a medical image, the visual encoder extracts a sequence of image features I. The text decoder, which can be either an RNN or a Transformer model, generates a sequence of words $\{w_1, w_2, ..., w_T\}$ to describe the medical image in an autoregressive manner. At each time step t, the decoder generates the next word w_t based on the previous words $\{w_1, w_2, ..., w_{t-1}\}$ and image features I. Assuming that the GT report is $\{w_1^*, w_2^*, ..., w_T^*\}$, the cross-entropy loss at each time step t is given by:

$$\mathcal{L}_{CE}(t) = -\log P(w_t^* \mid w_1^*, ..., w_{t-1}^*, I) \quad (1)$$

The total loss for the entire sequence is the sum of the losses over all time steps:

$$\mathcal{L}_{CE} = \sum_{t=1}^{T} \mathcal{L}_{CE}(t) = -\sum_{t=1}^{T} \log P(w_t^* \mid w_1^*, ..., w_{t-1}^*, I) \quad (2)$$

https://arxiv.org/html/2408.13988v1

Our Approach: MedCAM-OsteoCls

MedCAM-OsteoCls: Medical Context Aware Multimodal Classification of Knee Osteoarthritis

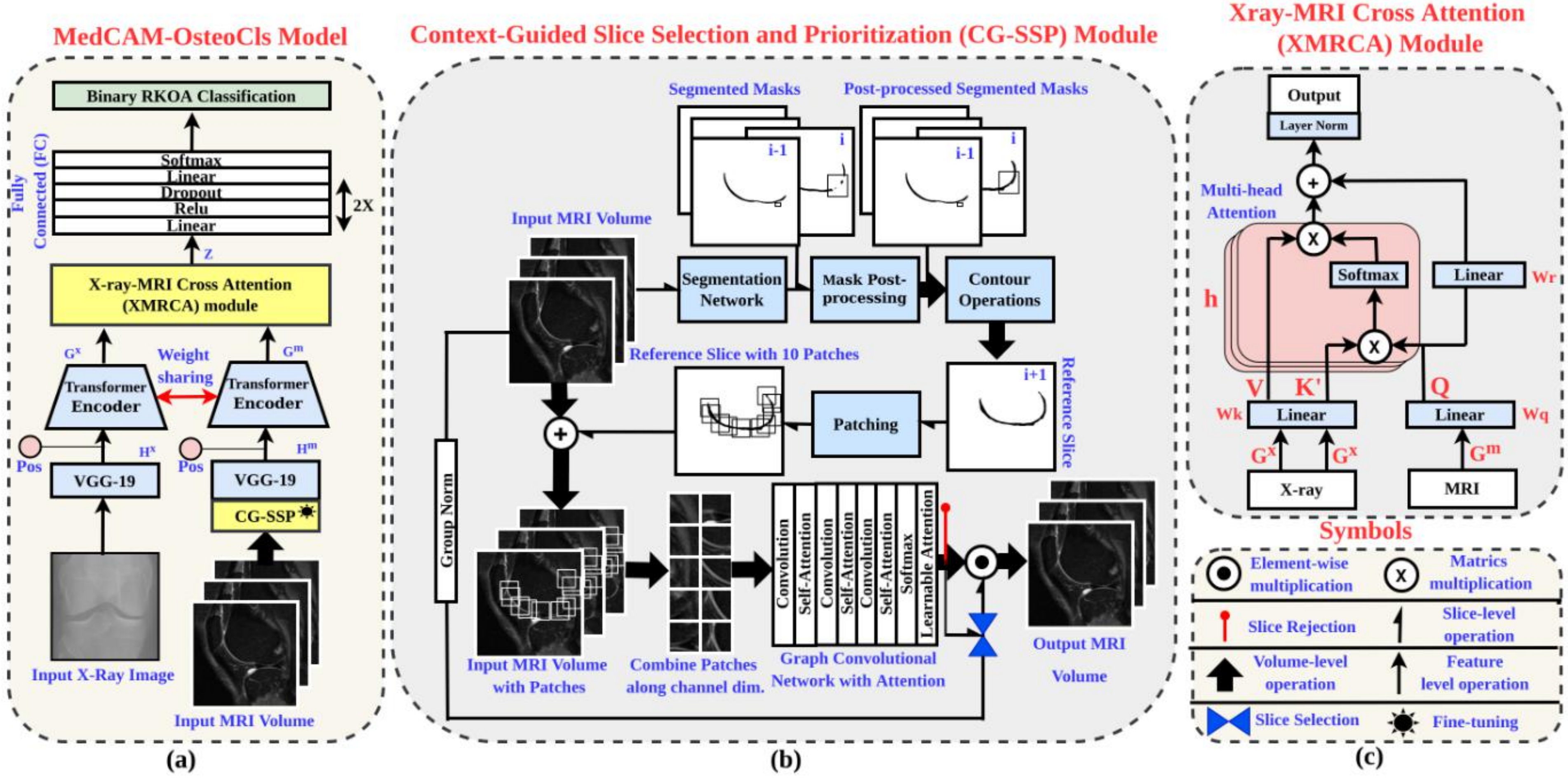


Fig. 1. Overall schematic of the proposed MedCAM-OsteoCls model with (a) VGG-19-TE +Fully Connected (FC) Network, (b) the CG-SSP module and (c) the XMRCA module.

Akshay Daydar

Research Scholar, IITG

Research Areas: Medical Imaging, Deep Learning, Biomechanics

https://adaydar.github.io/