COMP 448/548: Medical Image Analysis

Multiple-instance learning

Çiğdem Gündüz Demir
cgunduz@ku.edu.tr

1

Challenges

Build an image analysis model
Design an algorithm
Select parameters (if any)
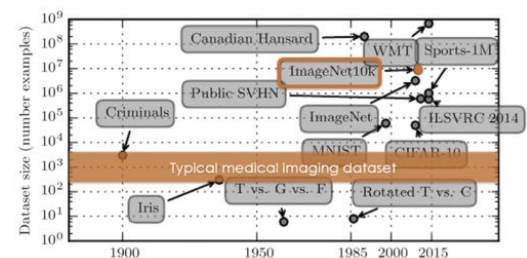**Train the model (if required)**

(revisited)

- Need of large annotated datasets to train deep models
  - **The more variety there is in the data, the larger the training dataset needs to be**

- Difficult to access to large high-quality annotated datasets
  - ImageNet is extremely powerful since it is huge and accurately annotated

- Imbalanced data problem

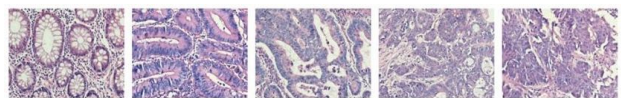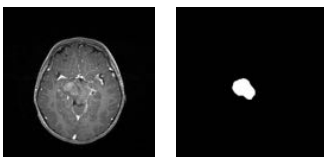www.quantib.com/blog/deep-learning-radiology-and-challenges-radiology-ai

2

1

Challenges

(revisited)

- Annotation is very challenging

- It needs medical expertise

- There might be inconsistencies in annotations

  - Sometimes there is no consensus among annotators
  Remember intra and inter-observer variability

  - There may exist hard-to-annotate image parts and incorrect annotations as a result

    - Due to noise and artifacts as well as due to the nature of problem

    - E.g., Detecting and marking all true positives in an image may not be possible or require too much effort
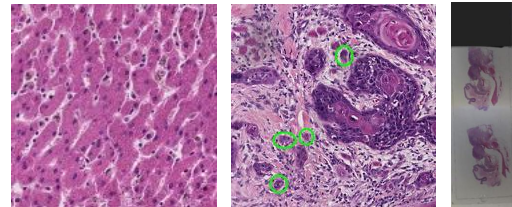
It is not like preparing a dataset for example for the application of pedestrian detection in a street, for which pedestrians can be marked by almost any person

3

To summarize …

- Although obtaining ground-truth local annotations (for patches or pixels) is costly, time-consuming or sometimes even impossible, *global labels are available for whole images*

These global labels do not apply to all patches or pixels inside a whole image
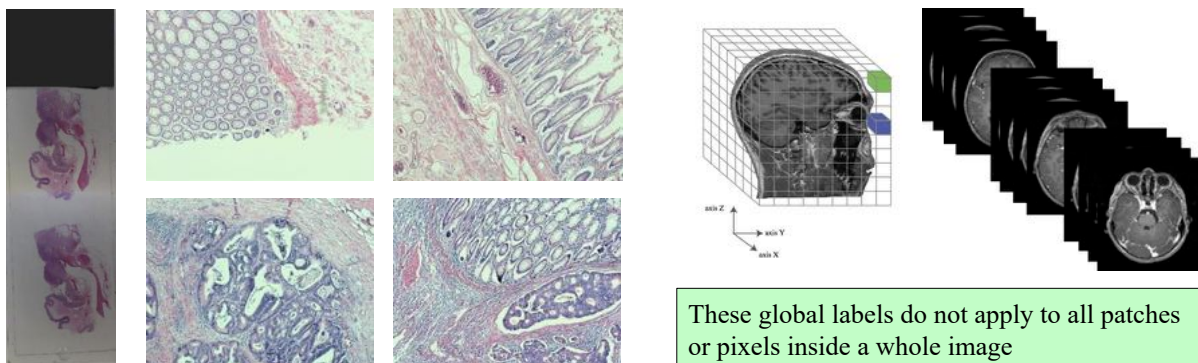
4

Multiple-instance learning

- Multiple-instance learning (MIL) is an extension of supervised learning to train learners using such weakly labeled data (i.e., only global labels are available)

- Training instances are arranged in sets (called _bags_) and a label is provided for an entire bag (not for every instance separately)

- In binary setting, a bag is labeled as

  - Positive if there is at least one positive instance in the bag

  - Negative if all instances in the bag are negative

> The main challenge is to cope with not knowing which instances in a positive bag are actually positive and which are negative

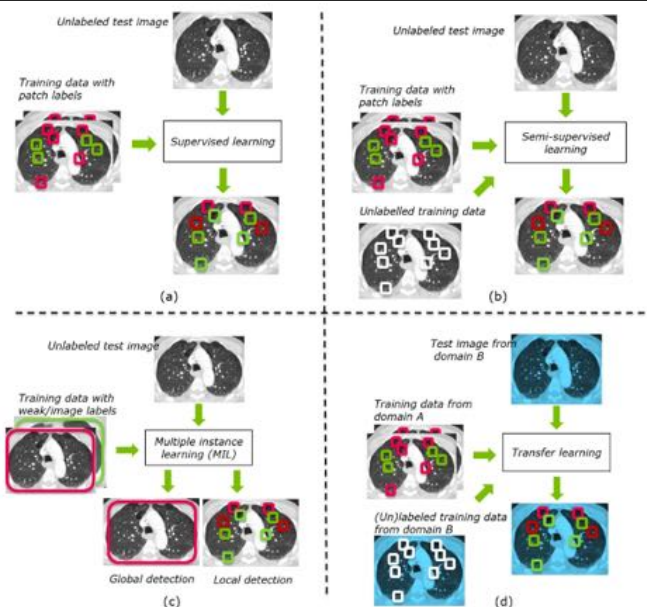Negative bags B_i^- Positive bags B_i^+
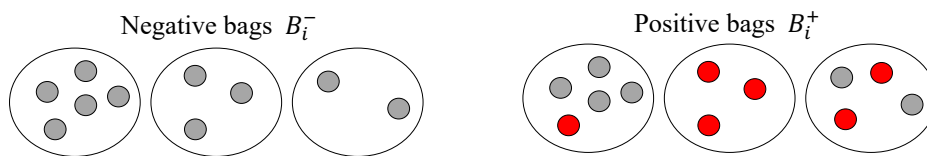
Fig. 1. Learning scenarios, illustrated by a task of classifying healthy (green) vs emphysema (red) tissue in chest CT images. Annotations are made for presentation purposes only and do not necessarily reflect ground truth. (a) Supervised learning, only healthy and abnormal patches are available. (b) Semi-supervised learning (Section 3), in addition to healthy and abnormal patches, unlabeled patches are available. (c) Multiple instance learning (Section 4). Labeled patches are not available, but subject-level labels (whether any abnormal patches are present) are. (d) Transfer learning (Section 5). Labeled patches are available, but for a different domain (here illustrated by different visual/characteristics) than the test set. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

In the binary setting of MIL, the focus could be

- **_Global detection_** in which an entire image (entire bag) is classified as either positive or negative and/or

- **_Local detection_** in which positives patches (instances) are localized on the entire image (entire bag)

Cheplygina et al., 2019. Not-so-supervised: A survey of semi-supervised, multi-instance, and transfer learning in medical image analysis.
https://www.sciencedirect.com/science/article/pii/S1361841 518307588

Collective approach

- In the standard setting, a bag is positive if it contains at least one positive instance

- Collective approach labels a bag as positive if it contains a "*certain*" instance combination

- Alternative assumptions:
  - Threshold-based assumption: A bag is positive if it contains at least a certain number of instances that belong to each required concept
  - Count-based assumption: A bag is positive if the number of instances from each required concept is between a minimum and a maximum number
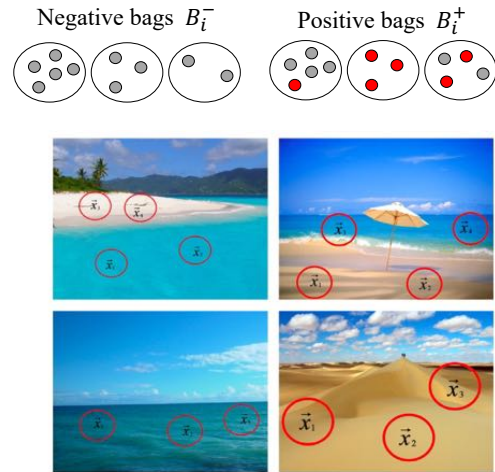
Negative bags B_i^- Positive bags B_i^+

Fig. 6. Classification of images into beach (top row) and non-beach (bottom row). See text.

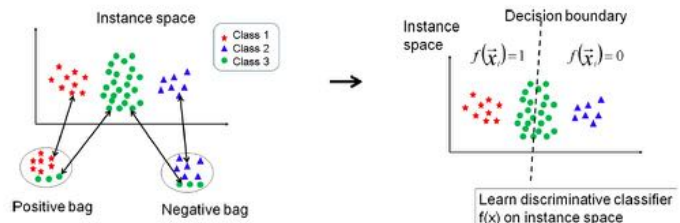*Amores, 2013. Multiple instance classification: Review, taxonomy and comparative study.*
https://www.sciencedirect.com/science/article/pii/S0004370213000581

Foulds and Frank, 2010. A review of multi-instance learning assumptions.
https://www.cs.waikato.ac.nz/ml/publications/2010/FouldsAndFrankMIreview.pdf

7

MIL paradigms

Instance-space paradigm

- This paradigm assumes that the discriminative information is at the instance-level

- It is local such that it employs the characteristics of individual instances, without considering more global characteristics of an entire bag

- It trains an instance-level classifier $f(x)$ to separate instances in positive bags from those in negative ones

- Given a new bag X, it obtains a bag-level classifier $F(X)$ by simply aggregating instance-level scores $f(x), x \in X$

Amores, 2013. Multiple instance classification: Review, taxonomy and comparative study.
https://www.sciencedirect.com/science/article/pii/S0004370213000581

8

MIL paradigms

Bag-space paradigm

Fig. 2. Illustration of the BS paradigm: training (a) and test (b). See text for an explanation.

- This paradigm assumes that the discriminative information is at the bag-level, and thus, it is global

- Each bag X is considered as a whole entity and a bag-level classifier $F(X)$ is trained to take a discriminative decision considering the global characteristics of the whole bag, instead of aggregating local instance-level decisions

- It employs non-vectorial learning
  - Defines a distance function $D(X,Y)$ that provides a metric to compare a pair of bags X and Y
  - Uses this distance function in any standard distance-based classifier such as kNN or any kernel-based classifier such as SVM
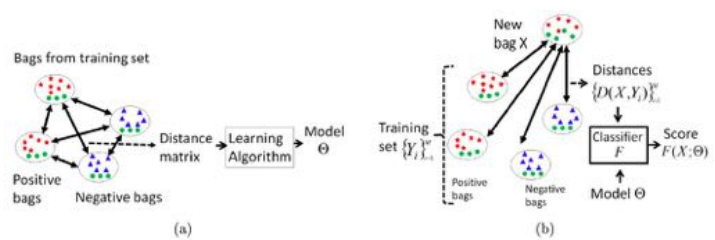
Amores, 2013. Multiple instance classification: Review, taxonomy and comparative study.
https://www.sciencedirect.com/science/article/pii/S0004370213000581

9

MIL paradigms

Embedded-space paradigm

Fig. 3. Illustration of the ES paradigm: training (a) and test (b). See text.

- This paradigm also exploits global, bag-level information

- It maps an entire bag X to a single feature vector, which summarizes the global characteristics of the entire bag

- It then trains a discriminative classifier on this vectorial embedded space

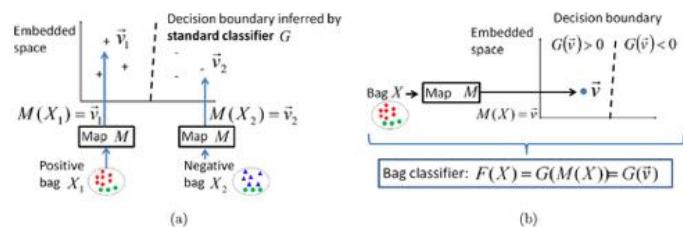- This indeed transforms the MIL problem into a standard supervised learning problem

Amores, 2013. Multiple instance classification: Review, taxonomy and comparative study.
https://www.sciencedirect.com/science/article/pii/S0004370213000581

10

MIL algorithms

- Many algorithms have been proposed for each of these paradigms

- Examples include
  - APR, mi-SVM, MI-SVM, Citation-kNN, mi-Graph, MILBoost, EM-DD, MILES, MILIS, DD-SVM, MILCNN, and more
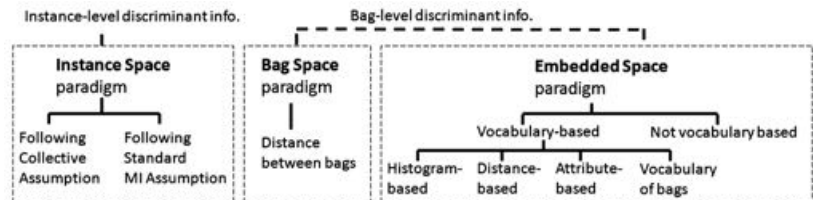

Instance-level discriminant info.   Bag-level discriminant info.

Instance Space paradigm	**Bag Space** paradigm	**Embedded Space** paradigm
Following Collective Assumption / Following Standard MI Assumption	Distance between bags	Vocabulary-based / Not vocabulary based — Histogram-based, Distance-based, Attribute-based, Vocabulary of bags

Fig. 4. Proposed taxonomy of MIC methods.

Amores, 2013. Multiple instance classification: Review, taxonomy and comparative study.
https://www.sciencedirect.com/science/article/pii/S0004370213000581

Foulds and Frank, 2010. A review of multi-instance learning assumptions.
https://www.cs.waikato.ac.nz/ml/publications/2010/FouldsAndFrankMIreview.pdf

Quellec et al., 2017. Multiple-instance learning for medical image and video analysis.
https://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=7812612

11

MIL in medical image analysis

Please check the following survey papers for more about multiple instance learning in medical image analysis

Quellec et al., Multiple-instance learning for medical image and video analysis, IEEE Reviews in Biomedical Eng, 2017.
https://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=7812612.

Cheplygina et al., Not-so-supervised: A survey of semi-supervised, multi-instance, and transfer learning in medical image analysis, Medical Image Analysis, 2019.
https://www.sciencedirect.com/science/article/pii/S1361841518307588

12

Example: MIL in digital mammography

- A bag is defined for each mammogram which is adaptively partitioned into regions (each corresponds to an instance)

- Each bag is labeled as positive if the corresponding mammogram contains an abnormal lesion (no region-based labeling)

- Each region (instance) is quantified with a set of features
  - Extracted using mass and microcalcification detection algorithms
  - Well known texture features (those defined on gray-level co-occurrence matrices, gray-level run-length matrices, and local binary patterns)

- Various MIL algorithms are used to classify a mammogram as well as to identify the most abnormal region(s) in the mammogram
  - Diverse density (DD), axis-parallel rectangles (APR), multiple-instance support vector machines (mi-SVM and MI-SVM), and MILBoost
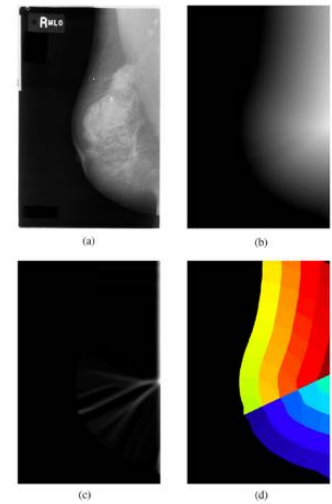

Fig. 1. Breast segmentation and region definition (a) original optical density image (b) distance transform (c) distance transform ridges (d) region segmentation.

Quellec et al., 2016. Multiple-instance learning for anomaly detection in digital mammography.
https://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=7390250

13

Example: MIL in CT scans

- A bag is defined for each CT scan

- Instances are patches extracted from CT images
  - Intensity patches are extracted from regions with high probability of containing lesions according to an atlas

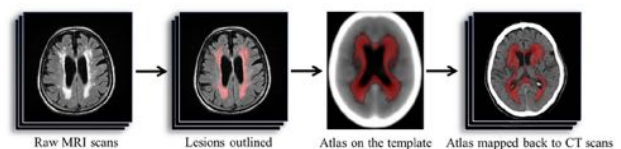- Boosting based MIL, called *MIS-Boost*, is used for the identification of cerebral small vessel disease

Fig. 2. The process of atlas construction and mapping back. The red regions are the ROIs for patch extraction.

Chen et al., 2015. Identification of cerebral small vessel disease using multiple instance learning.
https://link.springer.com/content/pdf/10.1007%2F978-3-319-24553-9_64.pdf

Akbas et al., 2011. MIS-Boost: Multiple instance selection boosting.
https://arxiv.org/pdf/1109.2388.pdf

14

Example: MIL in pathology slides

- A bag is defined for each pathology slide

- Instances are manually identified ROIs on each slide

- In each ROI, individual cells are segmented (clustering followed by watershed)
  - Each cell is quantified with three features (perimeter, major-to-minor axis ratio, average gray-level intensity)
  - Each ROI (instance) is quantified with a set of statistical features (mean, standard deviation, median, and mode) calculated on the extracted cell features

- SVM-based MIL algorithm is used to classify breast lesions
  - It is based on implementing the large margin principle with different loss functions defined for positive and negative samples
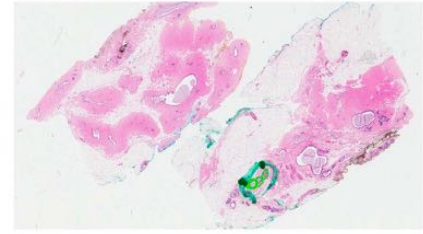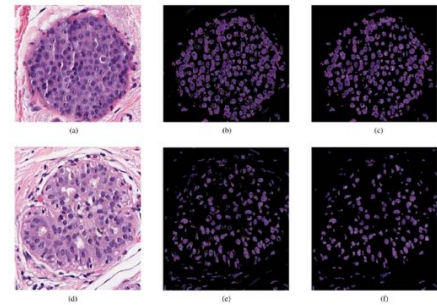

Fig. 1. Snapshot of the digitized scan of an H&E-stained specimen with three ROIs.

Dundar et al., 2011. Computerized classification of intraductal breast lesions using histopathological images.
https://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=5706360

15

Example: MIL and deep learning in pathology slides

- A bag is defined for a histopathological image

- Instances are pixels in the image

- Labels are available at the image-level (no pixel labels are available)

- A baseline FCN is trained to predict pixel labels

$$\mathcal{L}_{mil} = -\sum_i \left(\boldsymbol{I}(Y_i = 1) \log \widehat{Y}_i + \boldsymbol{I}(Y_i = 0) \log(1 - \widehat{Y}_i) \right)$$

$$\widehat{Y}_i = \left(\frac{1}{|X_i|} \sum_{k=1}^{|X_i|} \widehat{Y}_{ik}^r \right)^{1/r}$$
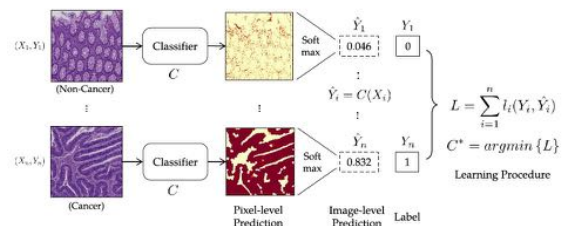

Fig. 1. Illustration of the learning procedure of a MIL algorithm. Our training dataset is denoted by $S = \{(X_i, Y_i), i = 1, 2, 3, \ldots, n\}$, where X_i indicates the ith input image, and $Y_i \in \{0, 1\}$ represents its corresponding manual label ($Y_i = 0$ refers to a non-cancer image and $Y_i = 1$ refers to a cancer image). Given an input image, a classifier C generates pixel-level predictions. Then, the image-level prediction $\widehat{Y}_i$ is computed from pixel-level predictions via a softmax function. Next, a loss between the ground truth Y_i and the image-level prediction $\widehat{Y}_i$ is computed for the ith input image, denoted by $l_i(Y_i, \widehat{Y}_i)$. Finally, an objective loss function L takes the sum of loss functions of all input images. The classifier C is learned by minimizing the objective loss function.

Jia et al., 2017. Constrained deep weak supervision for histopathology image segmentation
https://ieeexplore.ieee.org/stamp/stamp.jsp?arnumber=7971941

16

Example: MIL and deep learning in pathology slides

- **DWS-MIL (multiple instance learning framework with deep weak supervision)** aims to control and guide the learning process across multiple scales

- A side-output layer is defined for each scale and the FCN is trained to minimize the overall loss defined on all side-output layers

- Positive instances predicted by the FCN tend to progressively outgrow true cancerous regions. To address this problem, an area constraint term (penalty) is added for cancerous images
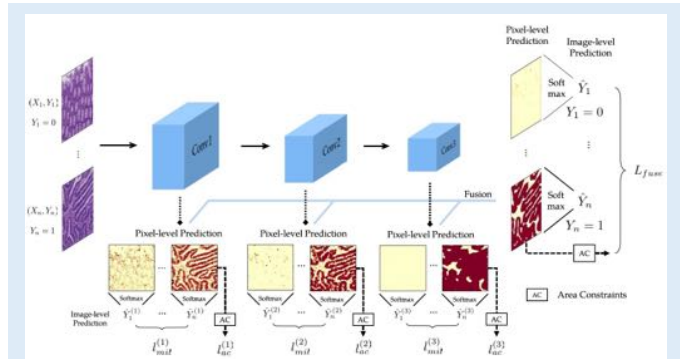

Fig. 3. Overview of our framework. Under the MIL setting, we adopt first three stages of the VGGNet and connect side-output layers with deep weak supervision under MIL. We also propose area constraints to regularize the size of predicted positive instances. To utilize the multi-scale predictions of individual layers, we merge side outputs via a weighted fusion layer. The overall model of equation (13) is trained via back-propagation using the stochastic gradient descent algorithm.

Jia et al., 2017. Constrained deep weak supervision for histopathology image segmentation
https://ieeexplore.ieee.org/stamp/stamp.jsp?arnumber=7971941

17

Example: MIL and deep learning in ultrasound images

- Hybrid use of a weakly annotated dataset and a smaller strongly annotated dataset

- Image-level loss

$$L_{ws}(I_i) = -\sum_{l \in \mathcal{L}} P_l^*(I_i) \log\big(P_l(I_i) \big)$$

where $P_l(I_i) = p_l(x_{MoI}), \forall l \in \mathcal{L}$
and image-level label set $\mathcal{L} = \{N, B, M\}$ for normal, benign, and malignant labels
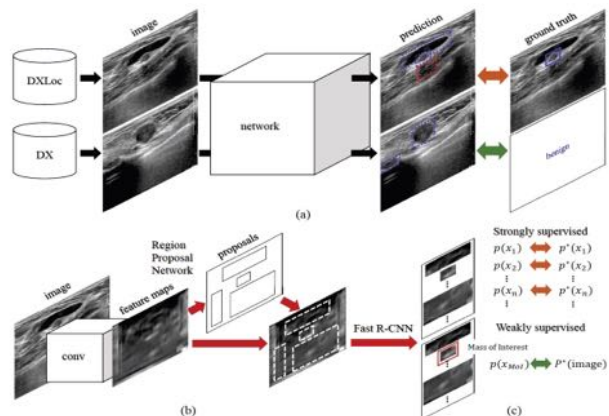x denotes a region and
x_{MoI} denotes a mass of interest

Fig. 2. Illustration of the proposed framework. (a) Images from two different data streams are forward-propagated into a shared network. (b) The Faster R-CNN [34] used for the "network" of (a). The network is composed of the region proposal network (RPN) and Fast R-CNN [35] with shared convolutional layers. This figure was previously presented in [34] and is reprinted in this paper for the description of the Faster R-CNN. We also note that the proposed method is a general framework; hence, other supervised approaches can also be adopted. (c) An image-level loss is used for images from DX, whereas region-level losses are used for images from DXLoc. Refer to Subsections II-B and II-C for details.

Shin et al., 2019. Joint weakly and semi-supervised deep learning for localization and classification of masses in breast ultrasound images.
https://arxiv.org/pdf/1710.03778.pdf

18

Example: MIL and deep learning in ultrasound images

$$L_{ws}(I_i) = -\sum_{l \in \mathcal{L}} P_l^*(I_i) \log\big(P_l(I_i) \big)$$

where $P_l(I_i) = p_l(x_{MoI}), \forall l \in \mathcal{L}$

It explores four criteria for mass of interest (MoI) selection

- Always select the most benign, malignant, discriminative, or abnormal region in the image

$$x_{MoI} = \arg\max_{x_n \in R(I_i)} p_B(x_n),$$

$$x_{MoI} = \arg\max_{x_n \in R(I_i)} p_M(x_n),$$

$$x_{MoI} = \arg\max_{x_n \in R(I_i)} \max_{l \in \{B,M\}} p_l(x_n),$$

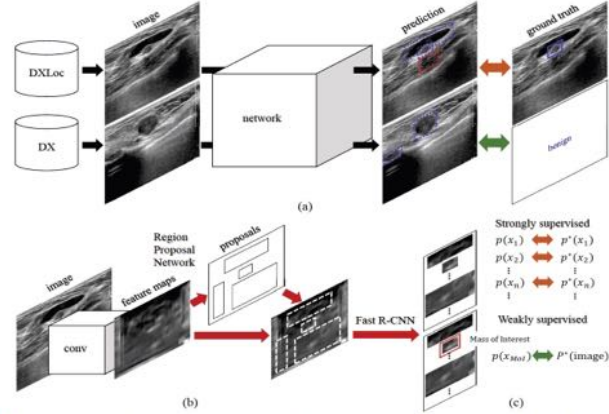$$x_{MoI} = \arg\min_{x_n \in R(I_i)} p_N(x_n),$$

Fig. 2. Illustration of the proposed framework. (a) Images from two different data streams are forward-propagated into a shared network. (b) The Faster R-CNN [34] used for the "network" of (a). The network is composed of the region proposal network (RPN) and Fast R-CNN [35] with shared convolutional layers. This figure was previously presented in [34] and is reprinted in this paper for the description of the Faster R-CNN. We also note that the proposed method is a general framework; hence, other supervised approaches can also be adopted. (c) An image-level loss is used for images from DX, whereas region-level losses are used for images from DXLoc. Refer to Subsections II-B and II-C for details.

Shin et al., 2019. Joint weakly and semi-supervised deep learning for localization and classification of masses in breast ultrasound images.
https://arxiv.org/pdf/1710.03778.pdf

19

Thank you!

20