Weak Supervision and Multiple-Instance Learning

Vladyslav Kolbasin

Senior Data Scientist at INSART
Consultant at Globallogic
Lecturer at CMAD dep. at NTU “KhPI”

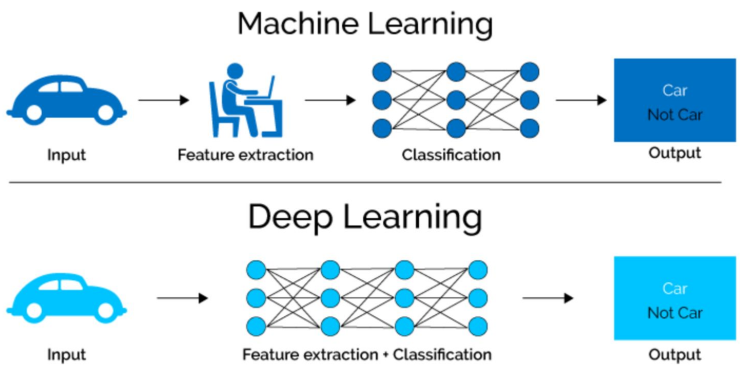
Deep Learning Approach

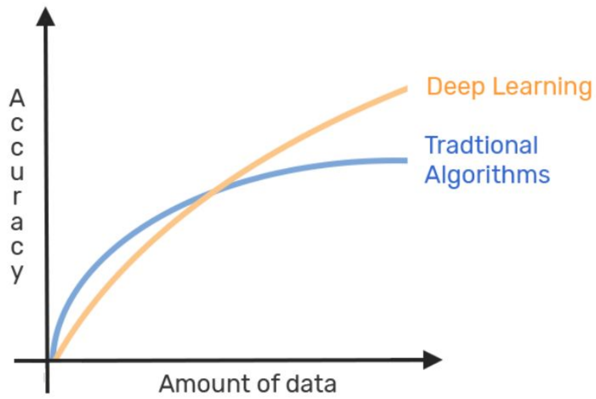
Pros:

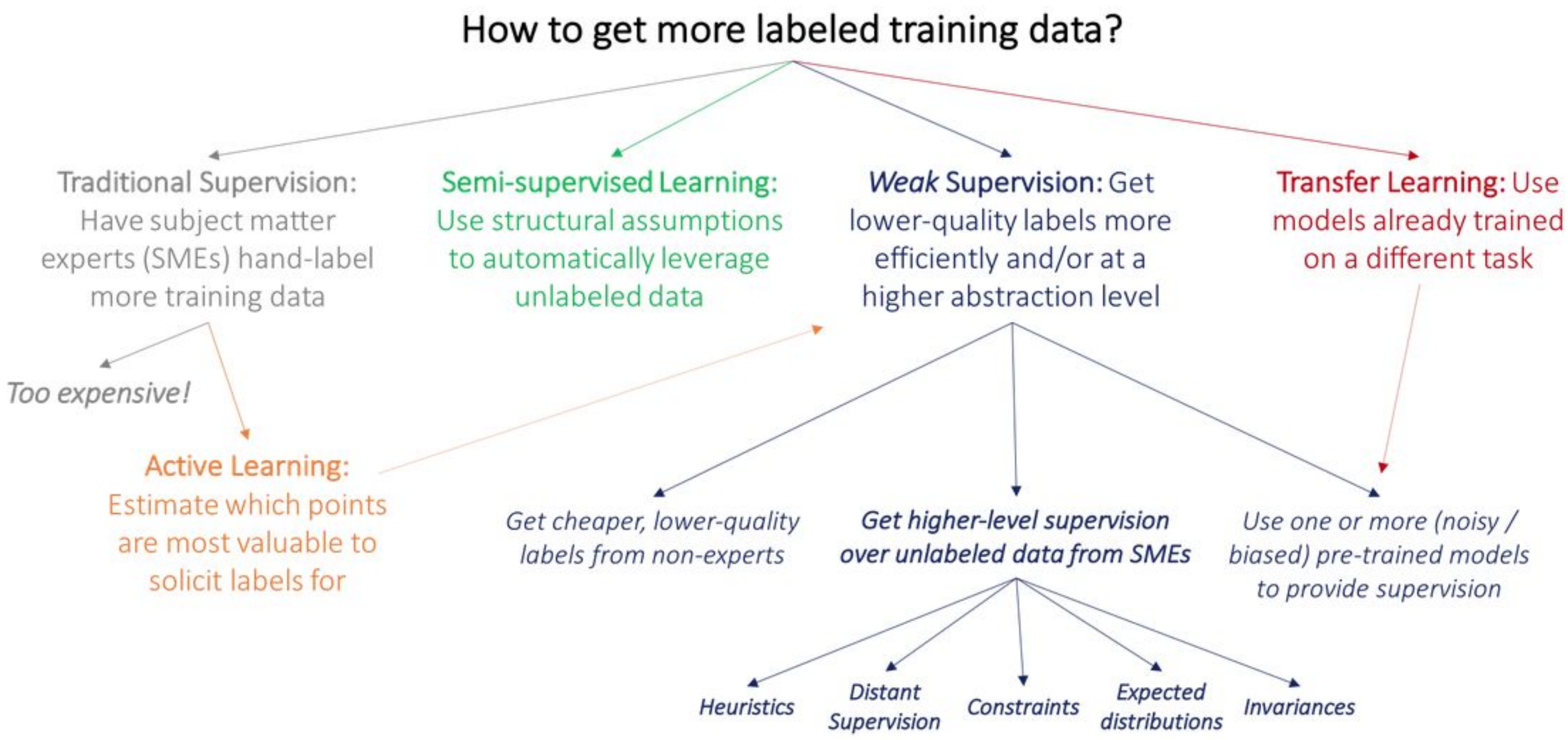
- Less focus on features crafting
- Transfer learning
- SOTA results

Cons:

- We heavily rely on massive sets of hand-labeled training data:
    - Labelling is expensive
    - Privacy concerns
    - Labelling requires domain experts
- Sometimes true labels are hidden
- Labelled data is static



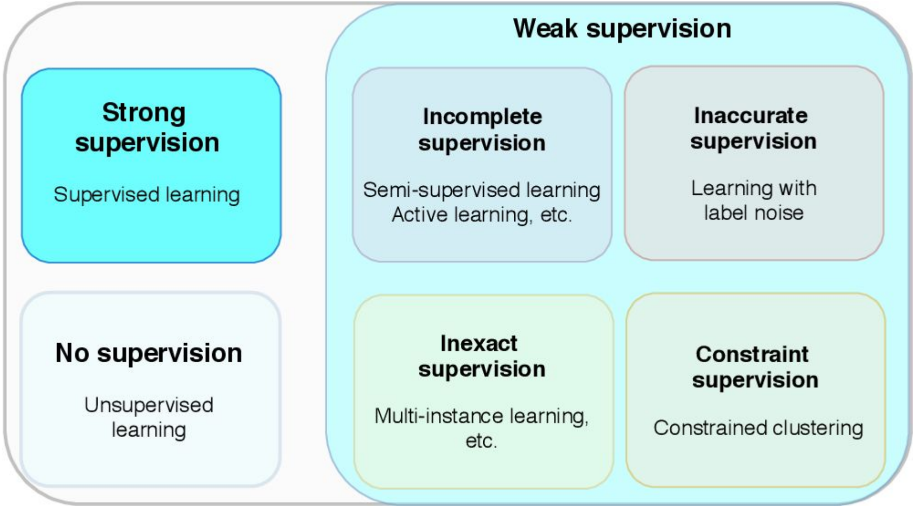




Credit: “Weak Supervision: The New Programming Paradigm for Machine Learning”
https://hazyresearch.github.io/snorkel/blog/ws_blog_post.html

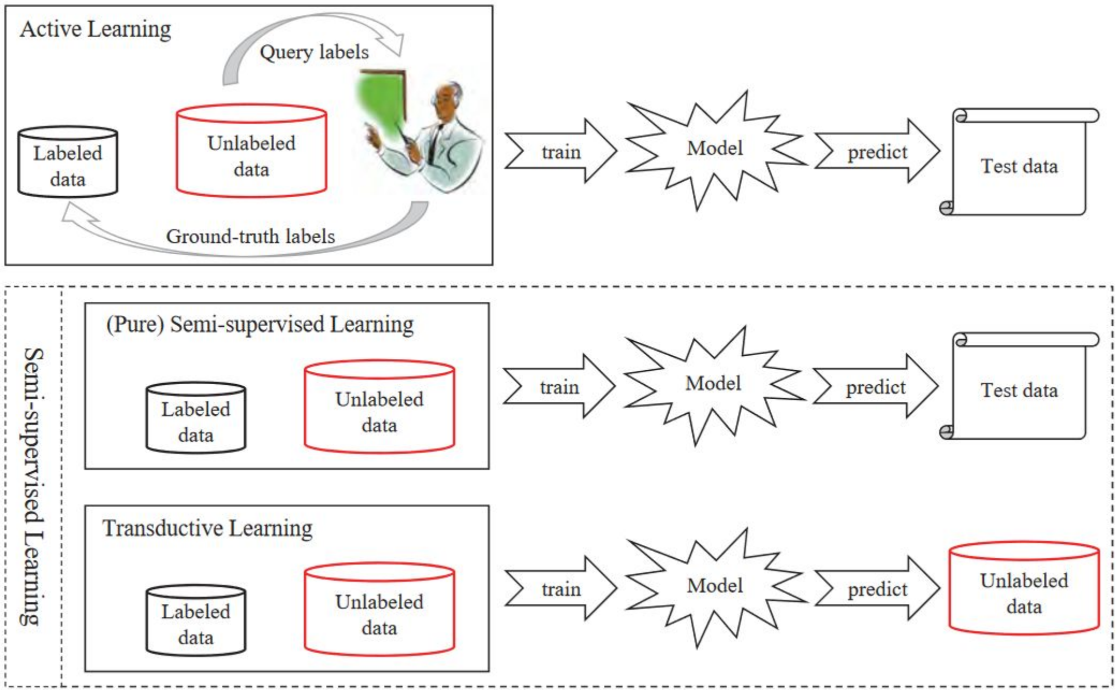
Weak Supervision

- There is no ground-truth labeled training set *for your business task*
- Instead we have one or more weak supervision sources
- Weak supervision source types:
  - **Incomplete supervision**. Only a subset of training data are given with labels
  - **Inaccurate supervision**. The given labels are not always ground-truth
  - **Inexact supervision.** The training data is given with only coarse-grained labels
- Usually we have a mix of all 3 types



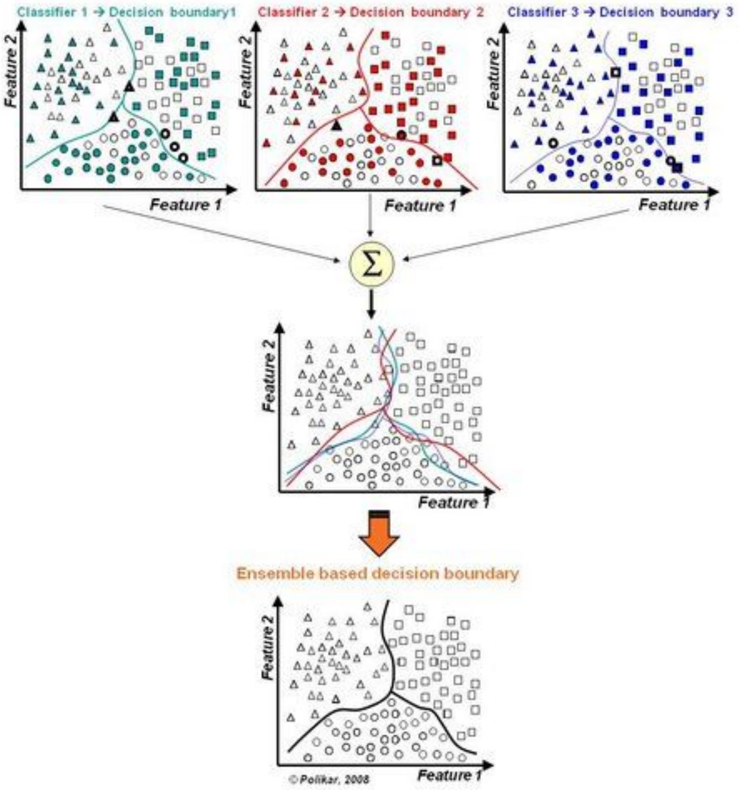
Weak Supervision Types

- **Incomplete supervision**. Only a subset of training data are given with labels
  - Active learning
  - Semi-supervised learning



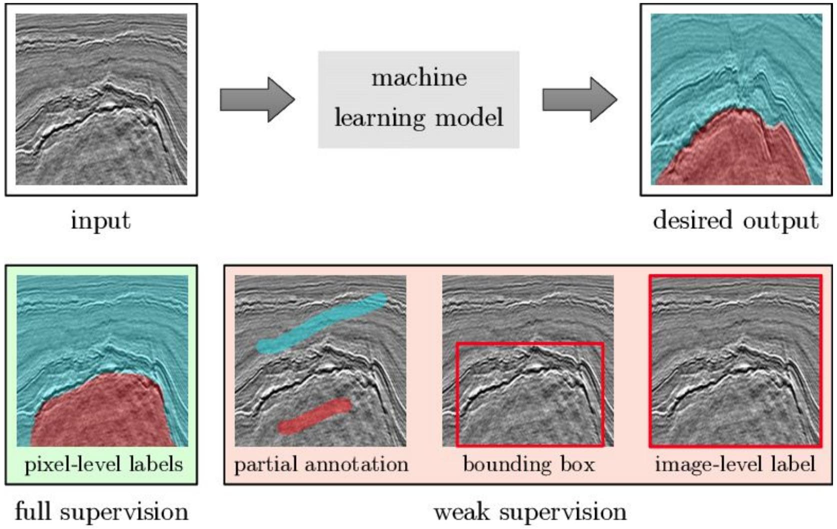
Weak Supervision Types

- **Inaccurate supervision.**
  The given labels are not always ground-truth
- We have ***weak labels***
- Reasons:
  - Labels are obtained via crowdsourcing
  - Labels are calculated using heuristic rules
  - Distant supervision
- Examples: Snorkel framework



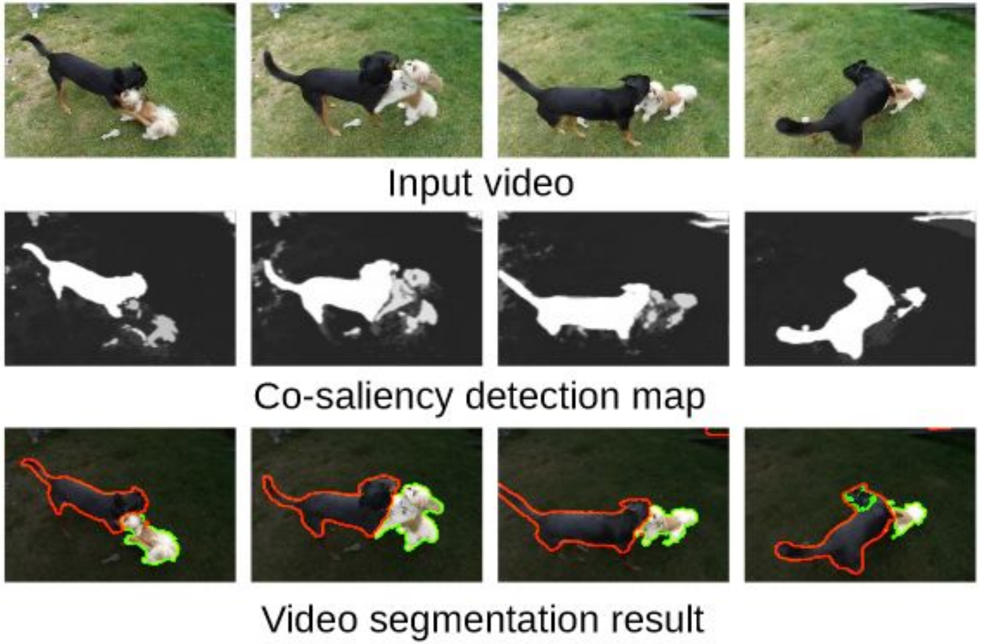
Weak Supervision Types

- **Inexact supervision.**
  The training data is given with only coarse-grained labels
  - Higher-level labels
  - Distributions
  - Multi-instance learning
- In other words:
  there are some latent variables



Weak Supervision. Co-saliency Detection

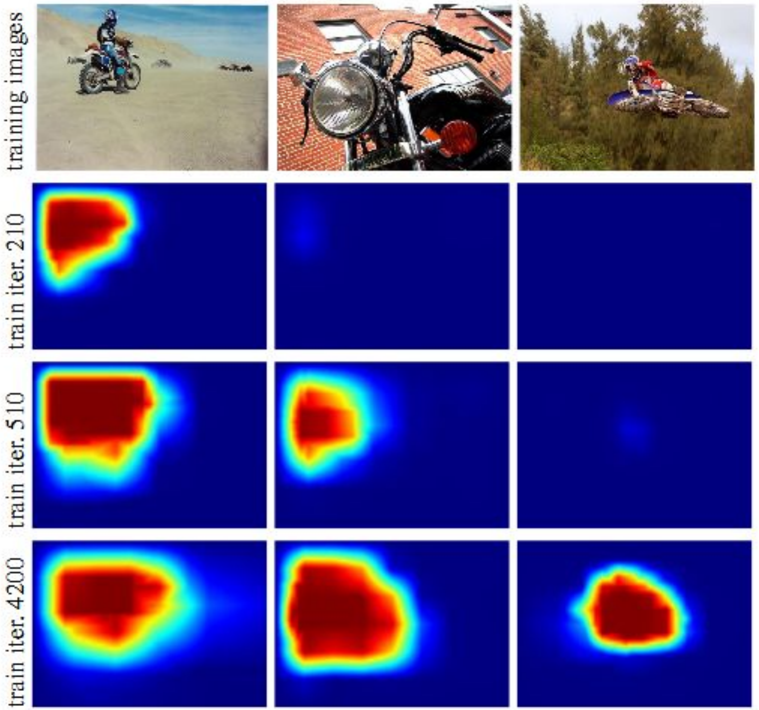
- Co-saliency detection refers to the discovery of common and salient foregrounds from two or more relevant images
- Weak Supervision Localization aims at learning object detectors to localize the corresponding objects by only using the image-level tags rather than the manually labeled bounding box annotations



https://arxiv.org/pdf/1604.07090.pdf

Inexact WS. Is object localization for free?

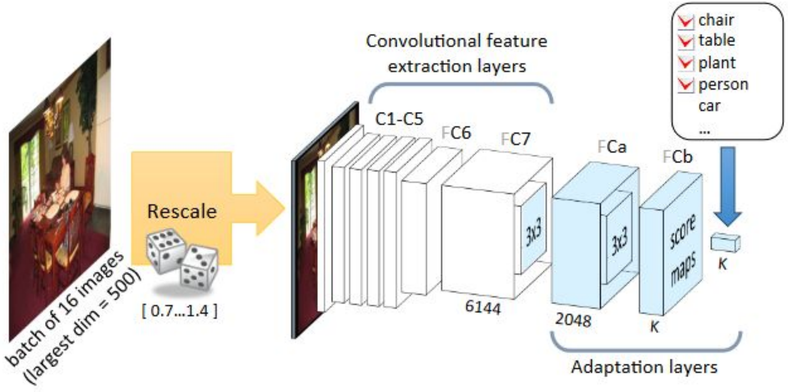
- **Goal**: Knowing only image labels,
  predict object localization
- **Results**:
  - outputs accurate image-level labels
  - predicts approximate locations (but not extents) of objects
  - performs comparably to its fully-supervised counter-parts using object bounding box annotation for training



https://www.di.ens.fr/~josef/publications/Oquab15.pdf

Inexact WS. Is object localization for free?

- Adaptation layers:
  - Create many FC and treat them as convolutions
    - Input 2048×1024 pixel => 58×26 output: score map for all classes for the different locations of the input 224×224 window with a stride of 32 pixels
  - Max-pooling layer that hypothesizes the possible location of the object in the image
  - Modify the cost function to learn from image-level super-vision



$$\ell(\, f_k(\mathbf{x})\, ,\, y_k\,) \;=\; \sum_k \log(1 + e^{-y_k f_k(\mathbf{x})})$$

Inexact WS. Is object localization for free?

	plane	bike	bird	boat	btl	bus	car	cat	chair	cow	table	dog	horse	moto	pers	plant	sheep	sofa	train	tv	mAP
I.MASKED POOL	89.0	76.9	**83.2**	68.3	39.8	**88.1**	62.2	90.2	47.1	83.5	40.2	88.5	**93.7**	83.9	84.6	44.2	80.6	**51.9**	**86.8**	64.1	72.3
J.WEAK SUP	90.3	77.4	81.4	**79.2**	41.1	87.8	66.4	**91.0**	47.3	**83.7**	55.1	**88.8**	93.6	85.2	87.4	43.5	**86.2**	50.8	**86.8**	66.5	74.5
K.CENTER PRED.	78.9	55.0	61.1	38.9	14.5	78.2	30.7	82.6	17.8	65.4	17.2	70.3	80.1	65.9	58.9	18.9	63.8	28.5	71.8	22.4	51.0
L.RCNN*	**92.0**	**80.8**	80.8	73.0	**49.9**	86.8	**77.7**	87.6	**50.4**	72.1	**57.6**	82.9	79.1	**89.8**	**88.1**	**56.1**	83.5	50.1	81.5	**76.6**	**74.8**

Table 3: Location prediction scores on the VOC12 validation set. Maximal responses are labeled as correct when they fall within a bounding box of the same class, and count as false negatives if the class was present but its location was not predicted. We then use the confidence values of the responses to generate precision-recall values.

Weak Supervision. Summary

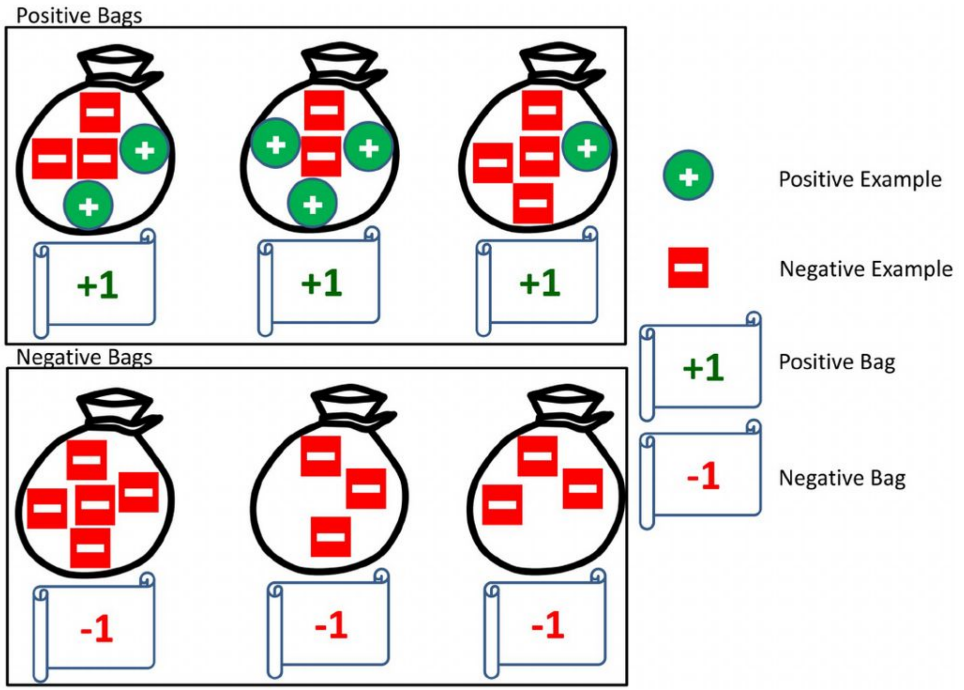
Motivation / Advantages:

- Reduce costs on data labelling/collection
- Deal with inherent ambiguity/human error
    - Generalize over set of weak supervision sources
- Include domain expertise into model
- Bonus: automatically discover latent information

Multiple-Instance Learning

Multiple Instance Learning. Definition

- Instead of receiving a dataset with individually labels, the learner receives a **set of labeled bags**, eac containing many instances
- A bag may be labeled:
  - **negative** if all the instances in it are negative
  - **positive** if there is at least one positive instance
- Individual labels of the instances contained in the bags are not provided



Multiple Instance Learning. Notation

- Training input:

Bags: $\{X_1, X_2, \ldots, X_n\}, X_i = \{x_{i1}, \ldots, x_{im}\}$

Bag labels: $\{y_1, \ldots, y_n\}, y_i \in \{0,1\}$

Instance labels: Y_{ij} are unknown during training

- Positive bag contains at least one positive instance:

$$y_i = max_j \; y_{ij}$$

Multiple Instance Learning. Typical Goals

1. Classify new and previously unseen bags as accurate as possible

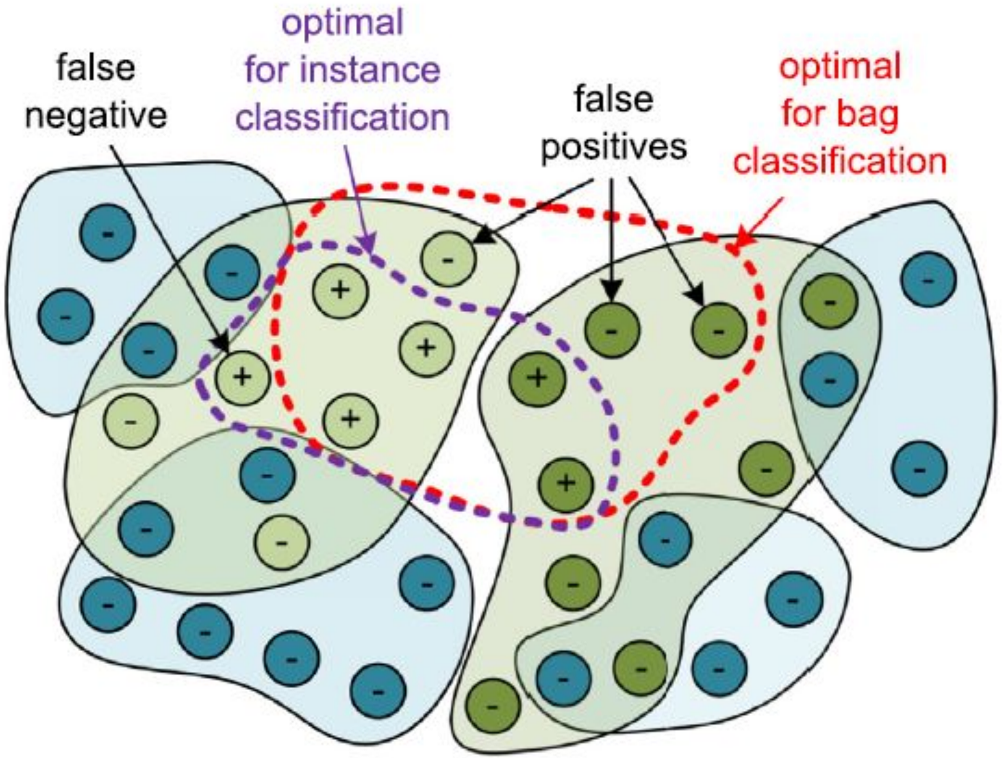
   $$f: \{X_i\} \rightarrow \{0,1\}$$

2. Sometimes classify each instance separately

   $$h: X \rightarrow \{0,1\}$$

3. Discover a “concept” that determines the positive class
   - Concepts are feature vectors that uniquely identify a positive class

Instance-level vs Bag-level Prediction

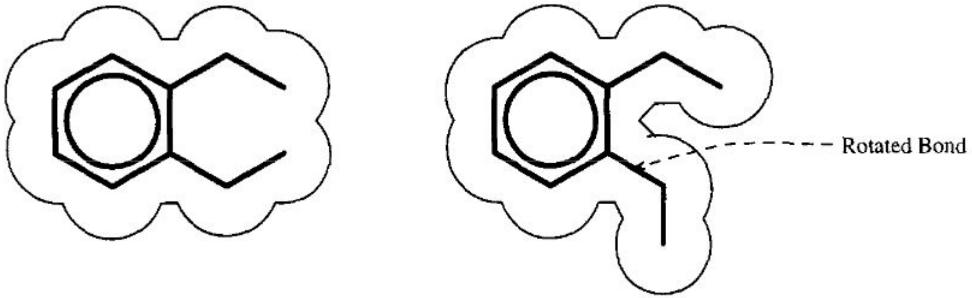


Where to use Multiple Instance Learning (MIL)?

Instances are naturally arranged in sets

- Many problems in biology, bioinformatics and chemistry

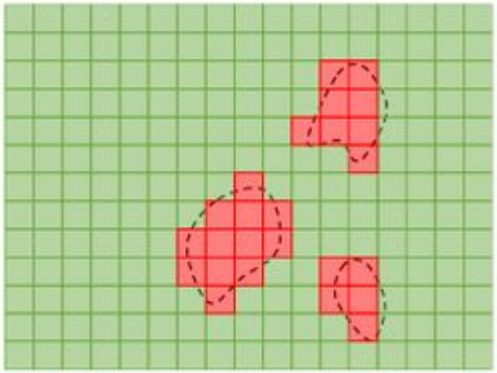
- Drug activity prediction problem.
  A molecule can take many conformations which can either produce, or not, a desired effect. Observing the effect of individual conformations is unfeasible.
  [Solving the multiple instance problem with axis-parallel rectangles]



- Compound object consists of several parts
  - Part-based models

To leverage weakly annotated data

- Computer-aided diagnosis algorithms can be trained with medical images for which only patient diagnoses are available instead of costly local annotations provided by an expert
- Object detectors trained on images collected with weak supervision
- Image & video classification
- Document classification
- Sound classification



(b) Detection — dense representation *(bag = image, instance = patch)*

(e) Event detection *(bag = video, instance = frame or video segment)*

MIL Algorithm Types. “Naive” Approach

- Copy bag labels to instances
- Train a regular classifier
- Combine all outcomes by simple combiner
  - E.g. max rule, averaging, majority vote, quantile / percentile

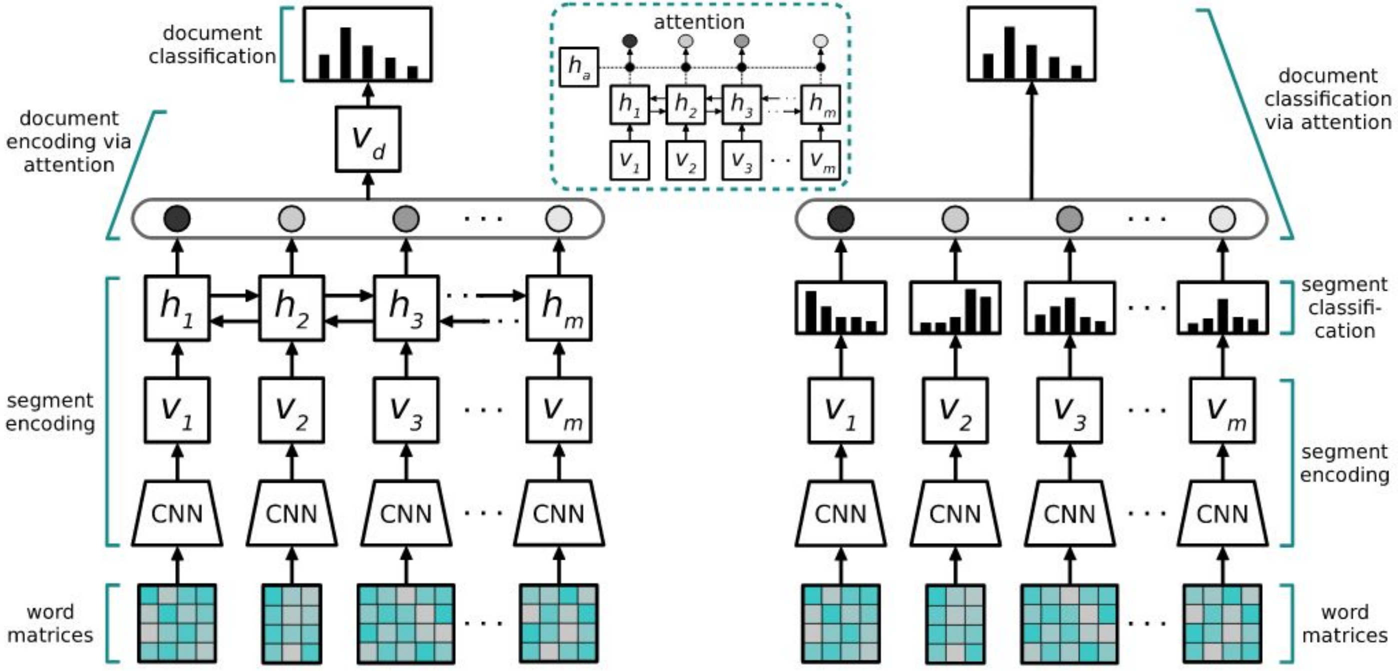
Can it work?

MIL Algorithm Types

- Instance-space algorithms - is better for interpretation
- Bag-space algorithm
- Embedding-space algorithms

is more accurate

1. Multiple Instance Learning Networks for Fine-Grained Sentiment Analysis

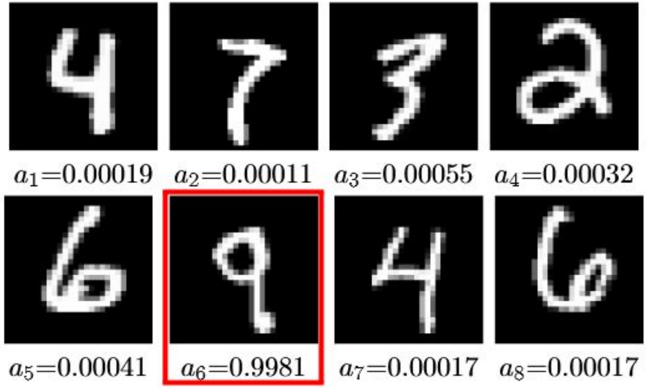


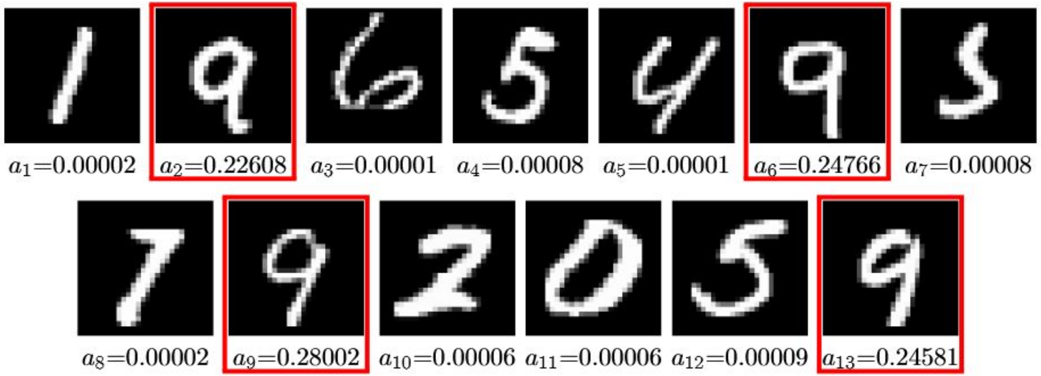
(a) HIERNET (b) MILNET

https://www.aclweb.org/anthology/Q18-1002

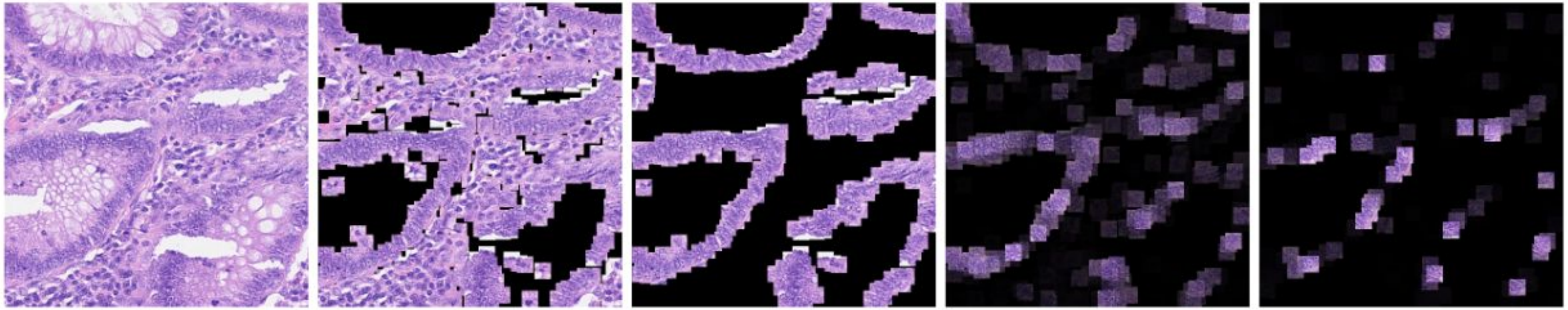
2. Attention-based Deep MIL

- Almost the same approach as (1), but for images
- Focus on interpretability as output - goal is to find key instances
    - Attention weights is a proxy for instance weights
- Change attention layer for sequentially independent instances
- More theoretical approach





https://arxiv.org/pdf/1802.04712.pdf



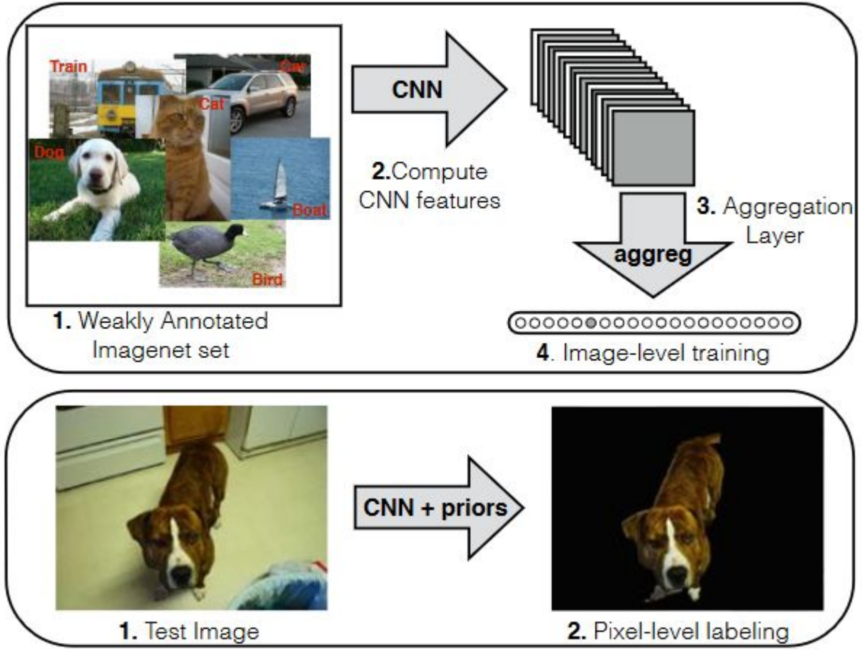
(a) (b) (c) (d) (e)

Figure 10. Colon cancer example 1: (a) H&E stained histology image. (b) 27×27 patches centered around all marked nuclei. (c) Ground truth: Patches that belong to the class epithelial. (d) Attention heatmap: Every patch from (b) multiplied by its attention weight. (e) Instance+max heatmap: Every patch from (b) multiplied by its score from the INSTANCE+max model. We rescaled the attention weights and instance scores using $a'_k = (a_k - \min(\mathbf{a}))/(\max(\mathbf{a}) - \min(\mathbf{a}))$.

(a) (b) (c) (d) (e)

3. From Image-level to Pixel-level Labeling with Convolutional Networks

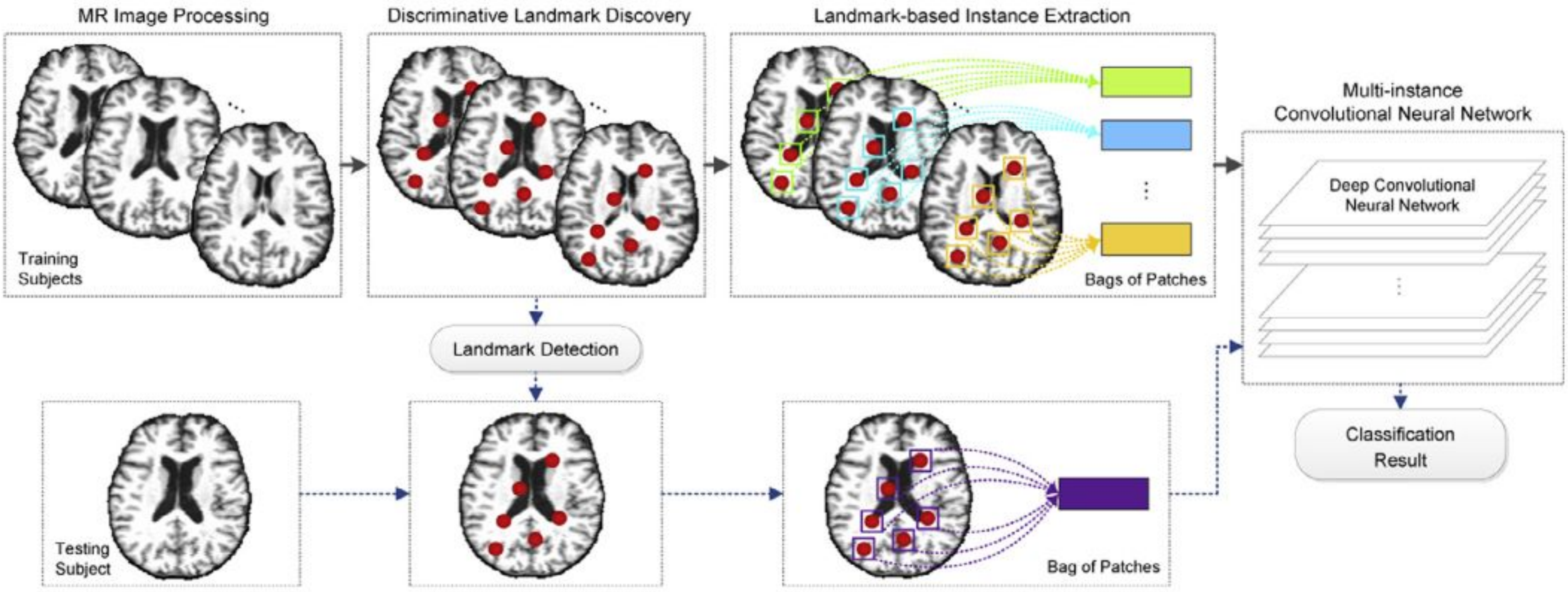
- Input: class labels for image
- Add prior knowledge about segmentation
- Output: Semantic segmentation



https://www.cv-foundation.org/openaccess/content_cvpr_2015/papers/Pinheiro_From_Image-Level_to_2015_CVPR_paper.pdf

Figure 4: **Inference results**. For each test image (left), we show the output assuming the image-level prior (center) and image-level and *SP-seg* smoothing prior (right).

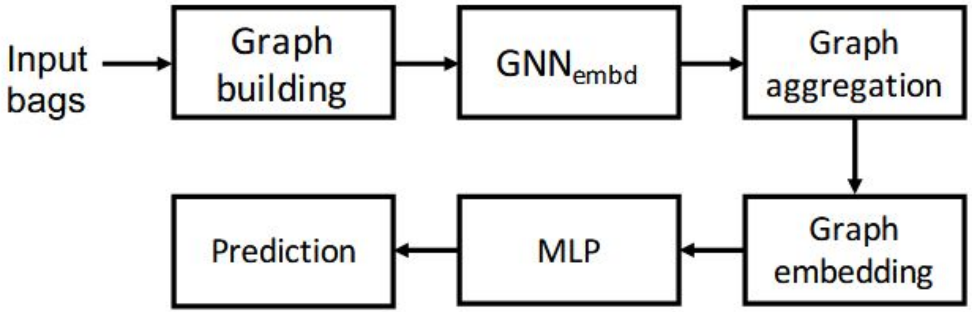
4. Landmark-based deep multi-instance learning for brain disease diagnosis



http://adni.loni.usc.edu/adni-publications/Landmark-based%20deep%20multi-instance%20learning%20for%20brain%20disease%20diagnosis.pdf

5. Multiple instance learning with graph neural networks

- ICML 2019
- Embedding-space algorithm
  - Instead of regarding instances in MIL bags as i.i.d samples, authors first convert each bag of instances into a graph
  - Then use an end-to-end GNN based network to learn the representation of the graph as the embedding of the bag



https://graphreason.github.io/papers/32.pdf

Take-away message

1) Weak supervision is a too overloaded term
2) It is a hot theme (even on AIUkraine2019 many reports connect to it)
3) Hope a lot of solutions will appear soon in this area
4) Multiple-instance learning is (not?) an exotic type of learning
5) For some domains MIL is very natural
6) Many standard problems can be reformulated in MIL formulation

Reference. Weak Supervision

- A Brief Introduction to Weakly Supervised Learning
  https://pdfs.semanticscholar.org/3adc/fd254b271bcc2fb7e2a62d750db17e6c2c08.pdf
- Weak Supervision: The New Programming Paradigm for Machine Learning
  https://hazyresearch.github.io/snorkel/blog/ws_blog_post.html

Reference. MIL

Surveys on MIL:

1. https://www.etsmtl.ca/Unites-de-recherche/LIVIA/Recherche-et-innovation/Publications/Publications-2017/mil_marc_2017.pdf
    a. https://github.com/macarbonneau/MILSurvey
    b. https://github.com/macarbonneau/MILSurvey/tree/master/Datasets
2. https://www.researchgate.net/publication/312173565_Multiple-Instance_Learning_for_Medical_Image_and_Video_Analysis
3. F. Herrera, S. Ventura, R. Bello, C. Cornelis, A. Zafra, D. Sánchez-Tarragó, S. Vluymans, Multiple Instance Learning: Foundation and Algorithms, Springer, 2016

Questions?

Vladyslav Kolbasin
https://www.linkedin.com/in/vladyslav-kolbasin-5070684/