IMAGE DATASET – CLASSIFICATION

$I_i, w_i \quad i = 1 \dots N$

I_i: Images — w_i: class labels

1. TRAIN A DL CLASSIFICATION – CNN ARCHITECTURE

$$\{I_i, w_i\}_{i \in training} \rightarrow [DL]$$

$$\{I_i\}_{i \in test} \rightarrow [DL] \rightarrow \{\tilde{w}_i\}_{i \in test}$$

$$\{w_i\}_{i \in test} \rightarrow (+, -, \times) \rightarrow \text{PERFORMANCE}$$

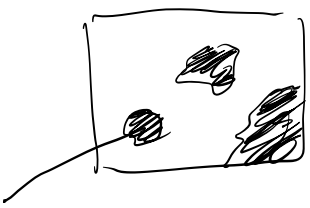
2. APPLY A DL INTERPRETABILITY TECHNIQUE {GRADCAM, Activation-Maximization, Saliency maps} IN ORDER TO OBTAIN AN INTERPRETABILITY HEATMAP

$$I_i \rightarrow [\text{DL Interpretability}] \rightarrow H_i \text{ Heatmap}$$

USE AN OTSU / MANUAL THRESHOLDING TO BINARIZE THE INTERPRETABILITY HEATMAP
('1' – REGIONS RELEVANT FOR THE PC CLASSIFIER
'0' – REGIONS NOT RELEVANT)

$H_i \rightarrow$ [OTSU] $\rightarrow$ BH_i

BINARY HEATMAP

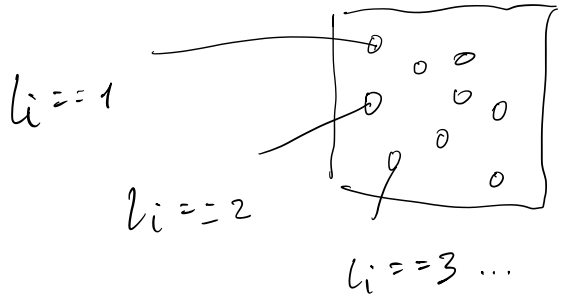


'1's: relevant regions

3. APPLY AN OBJECT SEGMENTATION METHOD TO I_i IN ORDER TO IDENTIFY RELEVANT STRUCTURES

$I_i \rightarrow$ [SEGMENTATION] $\rightarrow$ L_i

segmented/labelled Image



A TOTAL OF N_i OBJECTS ARE FOUND

4. EXTRACT REGIONAL ATTRIBUTES FROM THE SEGMENTED REGIONS:

- SHAPE
- AREA / SIZE
- TEXTURE
- ORIENTATION ...

→

OBJECT	FEATURES
1	$f_{11}\ f_{12} \cdots f_{1m}$
2	$f_{21}\ f_{22} \cdots f_{2m}$
3	
...	
N_i	$f_{N_i 1} \cdots f_{N_i m}$

A TOTAL OF m FEATURES ARE EXTRACTED

DATA MATRIX

$(N_i \times m)$

f_{ij} → feature

Object

5. ASSIGN A CLASS LABEL 'H' OR 'L' TO EACH OBJECT ACCORDING TO THE LOCATION OF THE CENTROID OF THE OBJECT IN THE BINARY HEATMAP:

OBJECTS LOCATED IN $BH == 1 \rightarrow$ 'H'

OBJECTS LOCATED IN $BH == 0 \rightarrow$ 'L'

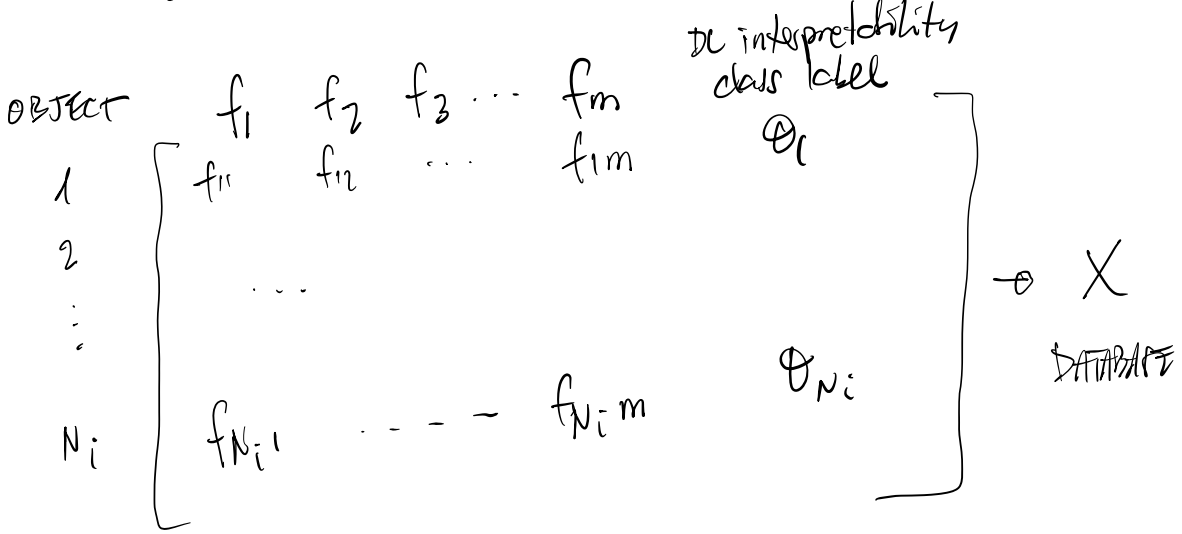
↓

INTERPRETABILITY CLASS VECTOR

$$\vec{\Theta} = \{\Theta_1, \ldots, \Theta_{N_i}\}$$

$$\Theta_i \in \{H, L\}$$

6. CONSTRUCT A CLASSIFICATION DATABASE WITH OBJECT FEATURES & OBJECT INTERPRETABILITY CLASS LABELS:



7. TRAIN A ~~DECISION TREE~~ WITH THE DATABASE X.

THE STRUCTURE OF THE TREE WILL REVEAL WHICH FEATURES ARE MORE RELEVANT IN ORDER TO DISTINGUISH BETWEEN RELEVANT & NON-RELEVANT REGIONS.