Ante-hoc explainability methods: the ProtoPNet architecture and its application on DBT images

ML-INFN

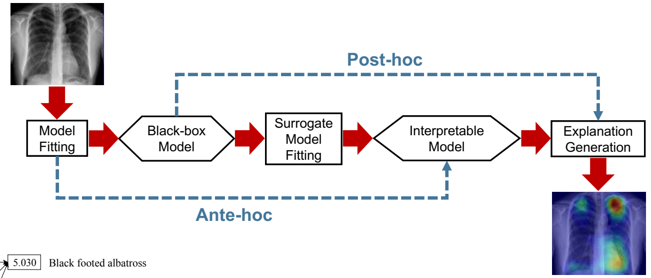
17/04/2023

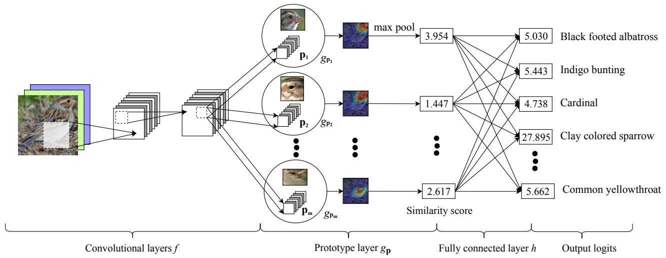
Andrea Berti

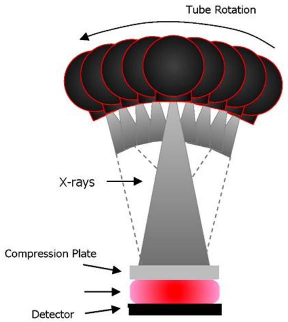
Istituto di Scienza e Tecnologie dell'Informazione "A. Faedo"
Consiglio Nazionale delle Ricerche

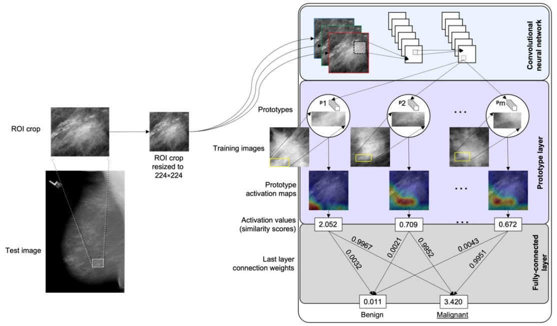
Overview

- The importance of explainability
- Ante-hoc explainability
- The ProtoPNet architecture
- Mammography and Digital Breast Tomosynthesis
- ProtoPNet on medical images









Zech, J. R., Badgeley, M. A., Liu, M., Costa, A. B., Titano, J. J., & Oermann, E. K. (2018). *Variable generalization performance of a deep learning model to detect pneumonia in chest radiographs: a cross-sectional study. PLoS medicine, 15(11), e1002683.*

Why explainability

Normal

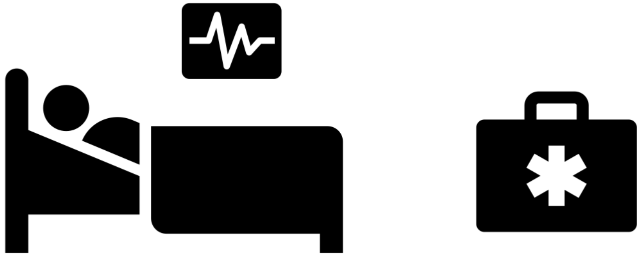
Pneumonia

Zech, J. R., Badgeley, M. A., Liu, M., Costa, A. B., Titano, J. J., & Oermann, E. K. (2018). *Variable generalization performance of a deep learning model to detect pneumonia in chest radiographs: a cross-sectional study. PLoS medicine, 15(11), e1002683.*

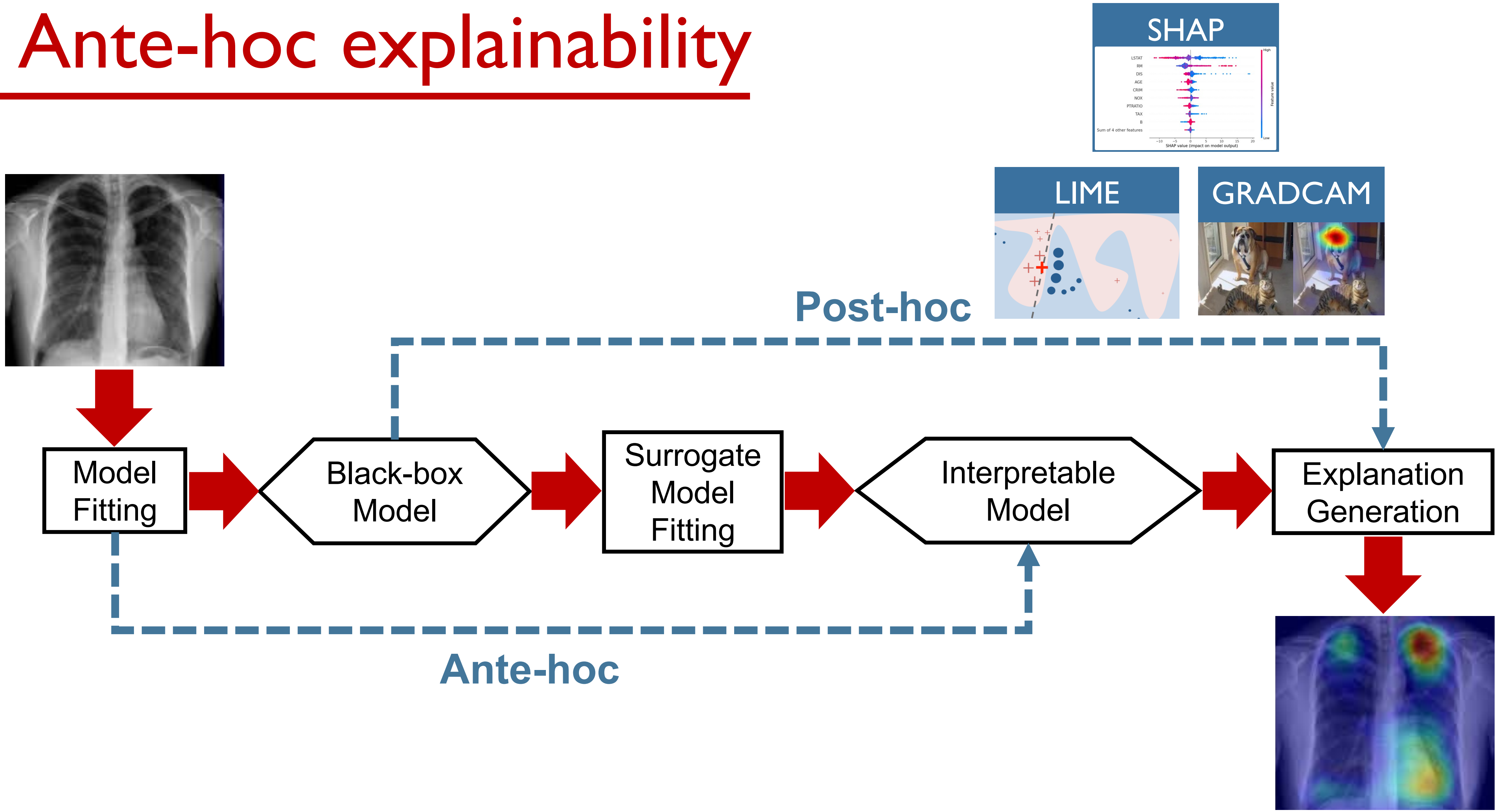
Why explainability

Normal

Pneumonia

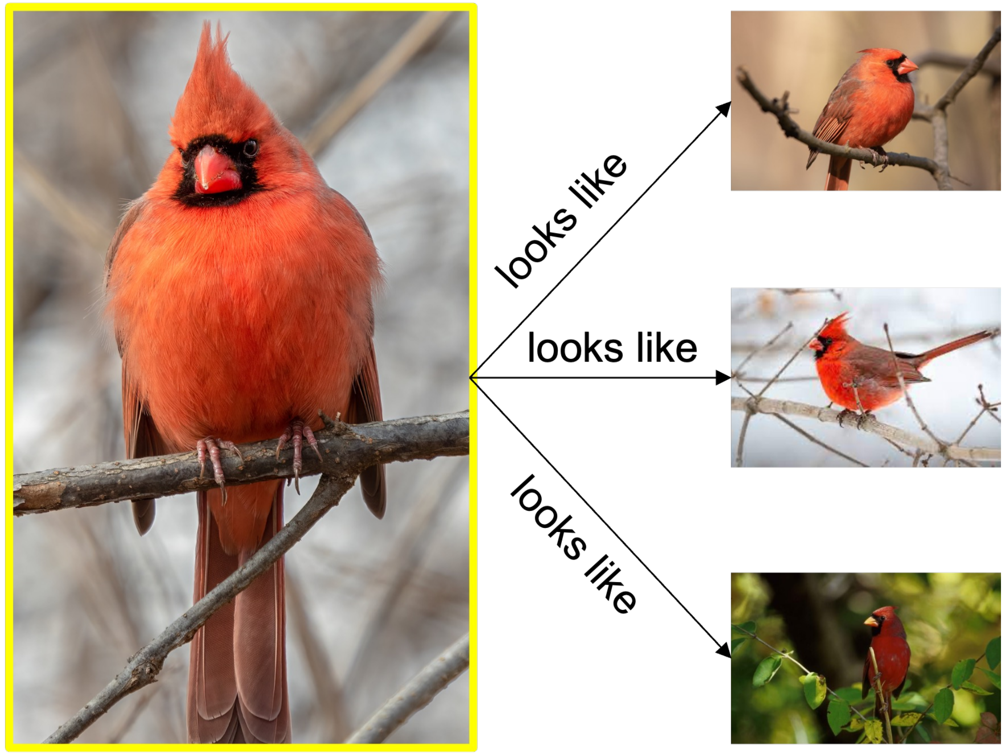


Ante-hoc explainability



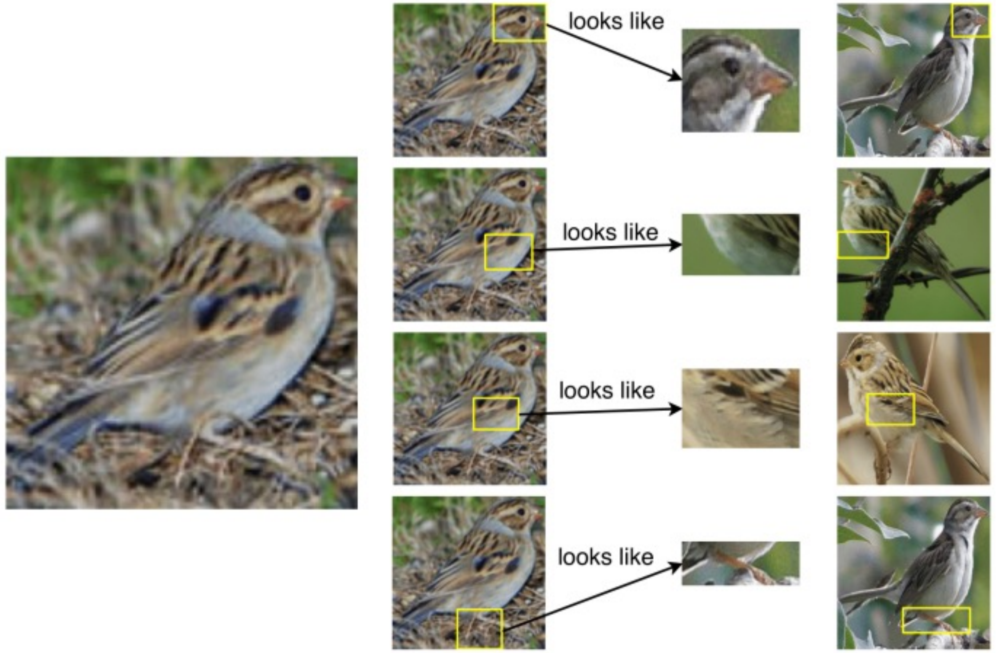
Case-based reasoning

Prototypical Learning



Northern Cardinal

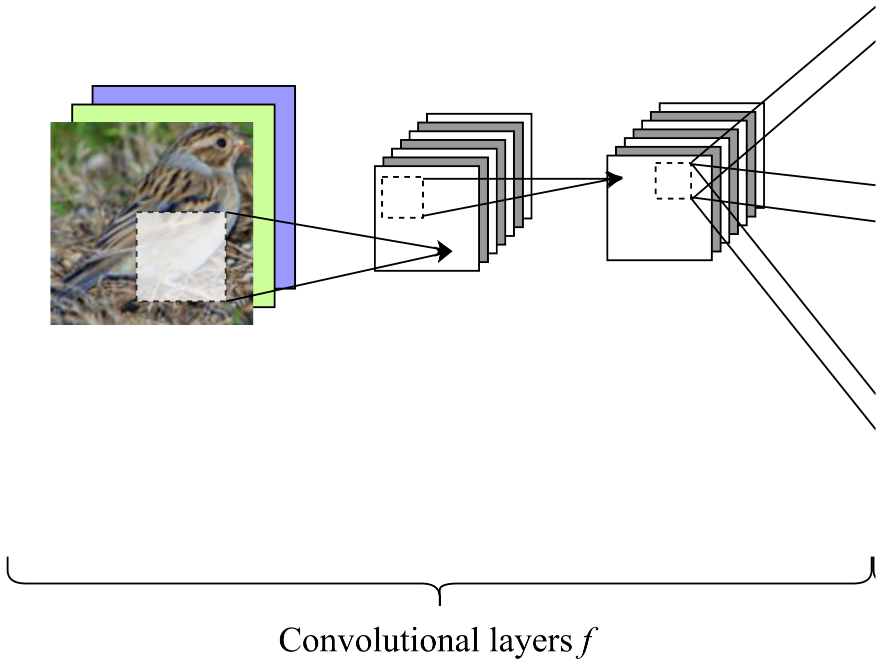
Prototypical Part Learning



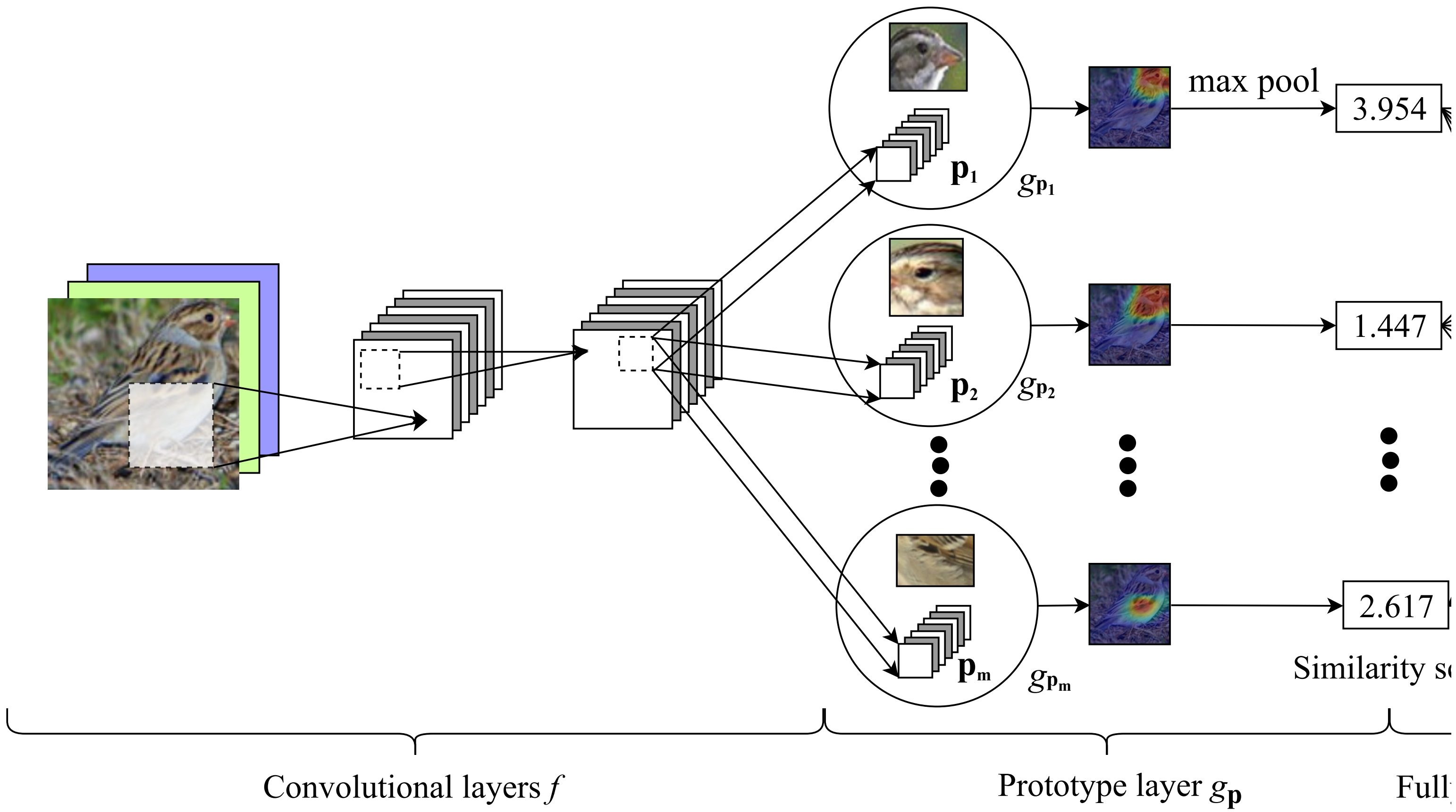
This looks like that: deep learning for interpretable image recognition

Chen, Li, et al. 2019, NeurIPS

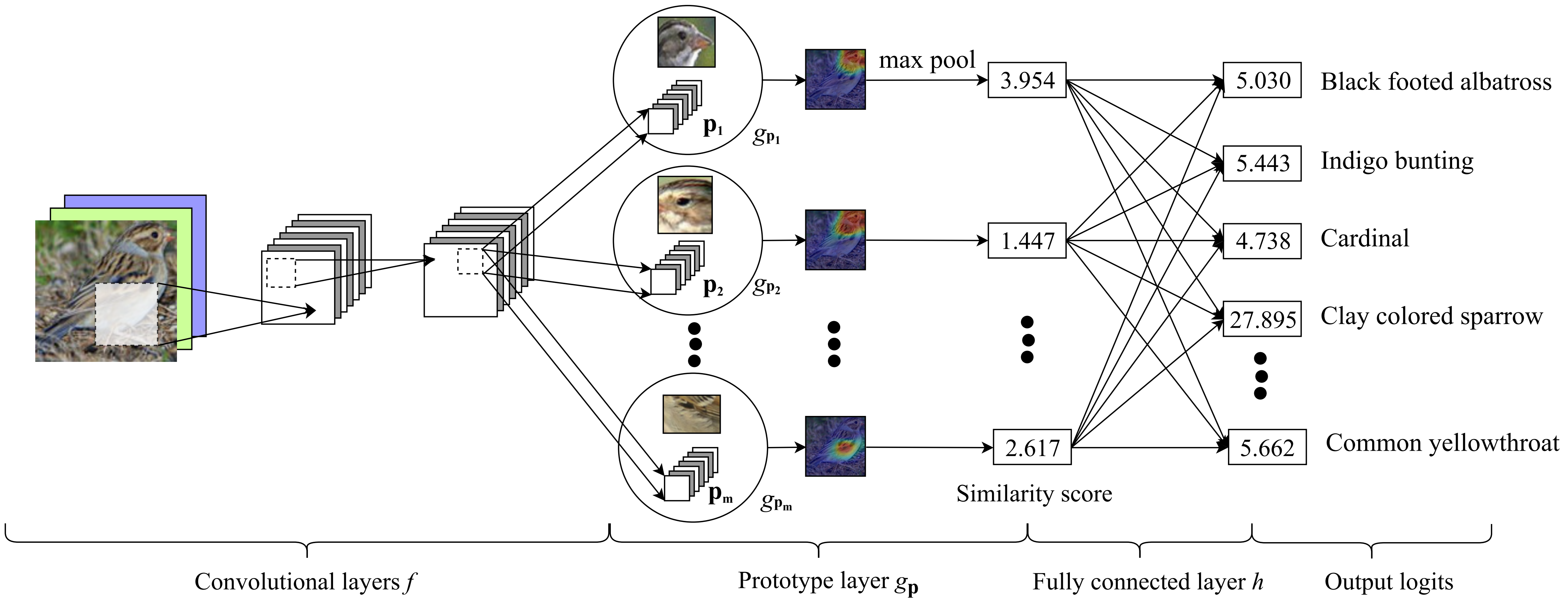
ProtoPNet architecture (inference)



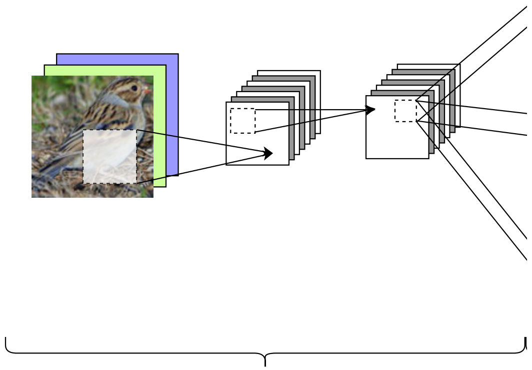
ProtoPNet architecture (inference)



ProtoPNet architecture (inference)



Convolutional layers

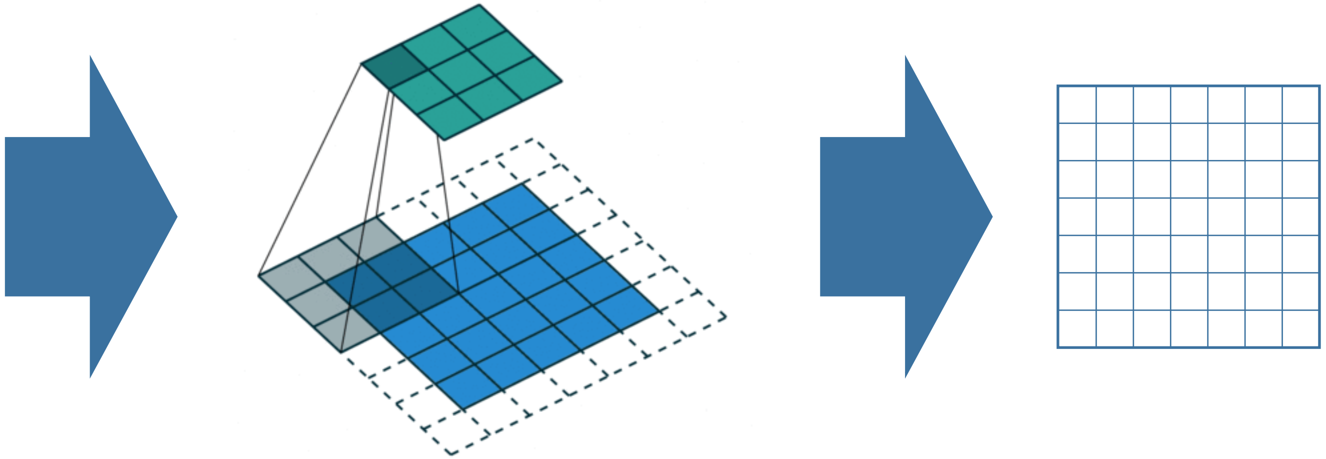


Convolutional layers f

- Feature extraction:
  - VGG
  - ResNet
  - DenseNet

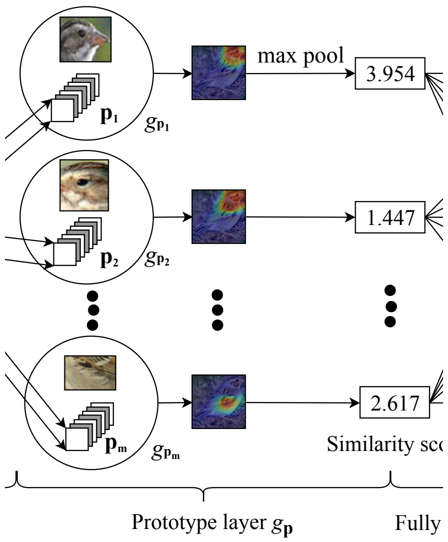
Pretrained on ImageNet

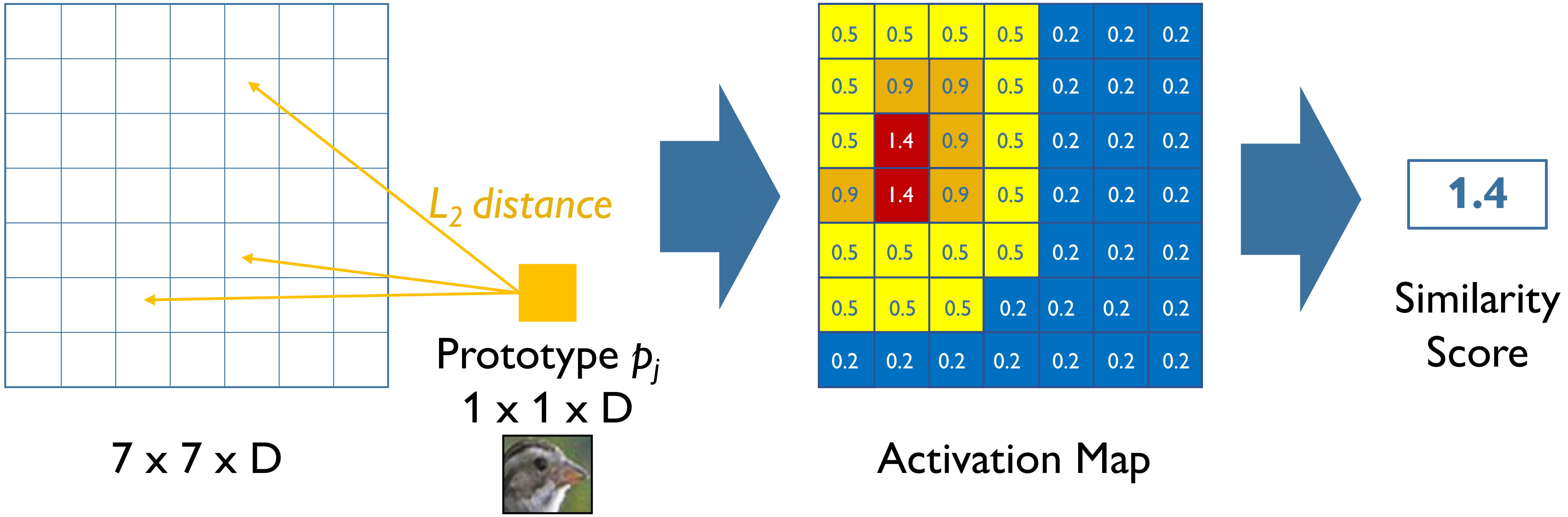
Original Image
224 x 224 x 3



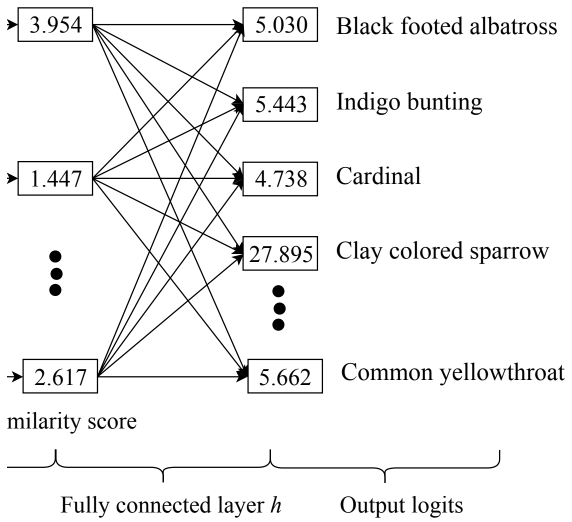
Latent Representation
7 x 7 x D

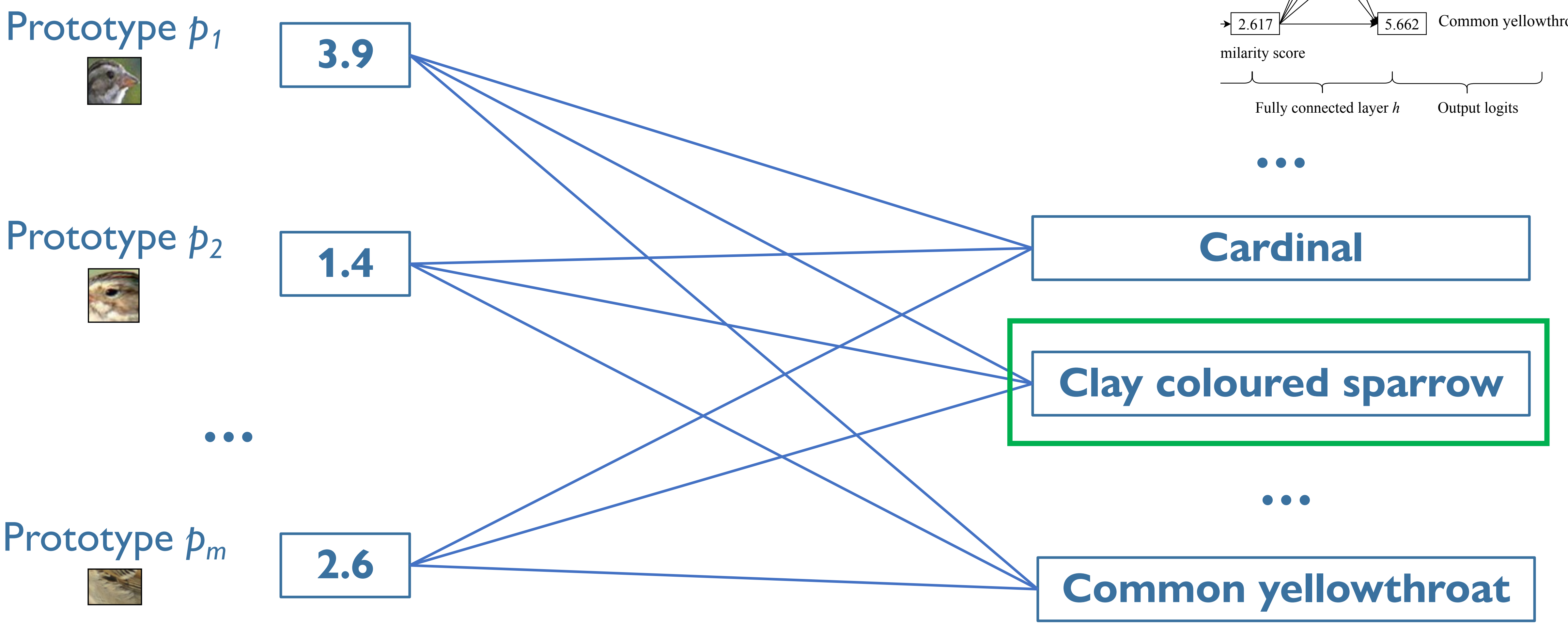
Prototype layer



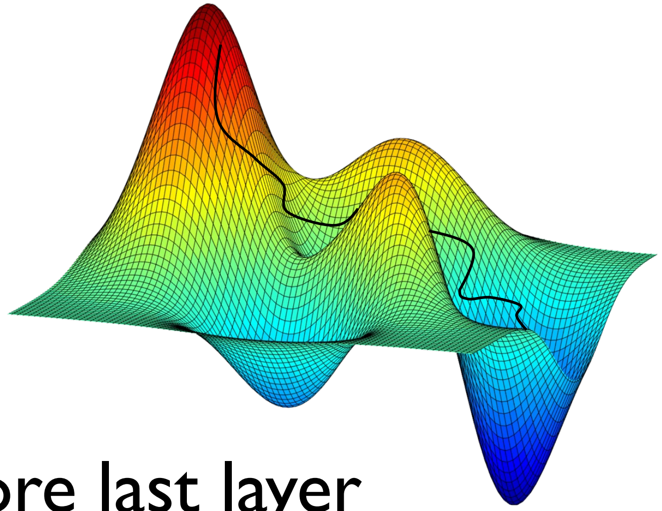


Fully connected layer



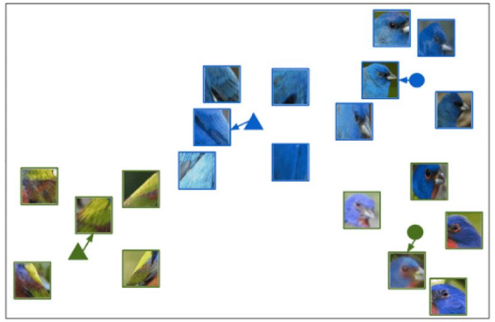


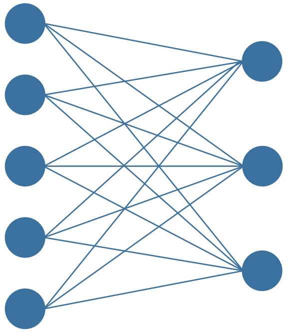
The training algorithm



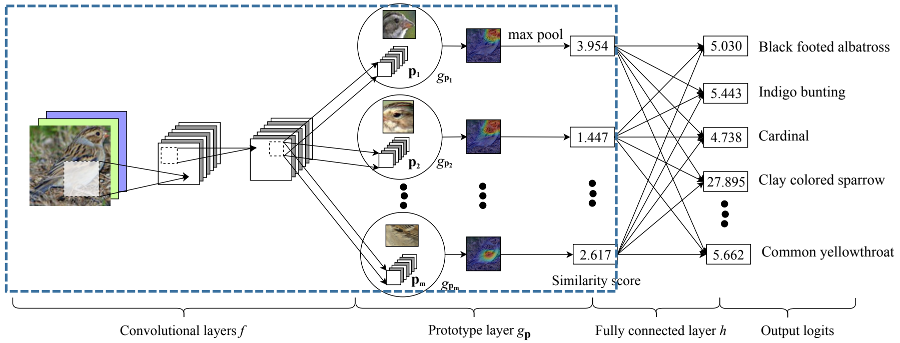
Three stages:

- Stage 1: Stochastic Gradient Descent of layers before last layer
- Stage 2: Projection of prototypes
- Stage 3: Optimization of last layer





The training algorithm (1/3)



Stage 1: SGD of layers before last layer

$$\min_{\mathbf{P}, w_{\text{conv}}} \frac{1}{n} \sum_{i=1}^{n} \text{CrsEnt}(h \circ g_{\mathbf{p}} \circ f(\mathbf{x_i}), \mathbf{y_i}) + \lambda_1 \text{Clst} + \lambda_2 \text{Sep}, \quad \text{where Clst and Sep are defined by}$$

$$\text{Clst} = \frac{1}{n} \sum_{i=1}^{n} \min_{j : \mathbf{p}_j \in \mathbf{P}_{y_i}} \min_{\mathbf{z} \in \text{patches}(f(\mathbf{x}_i))} \|\mathbf{z} - \mathbf{p}_j\|_2^2; \text{Sep} = -\frac{1}{n} \sum_{i=1}^{n} \min_{j : \mathbf{p}_j \notin \mathbf{P}_{y_i}} \min_{\mathbf{z} \in \text{patches}(f(\mathbf{x}_i))} \|\mathbf{z} - \mathbf{p}_j\|_2^2.$$

Clst: each training image has some latent patch close to, at least, one prototype of the same class

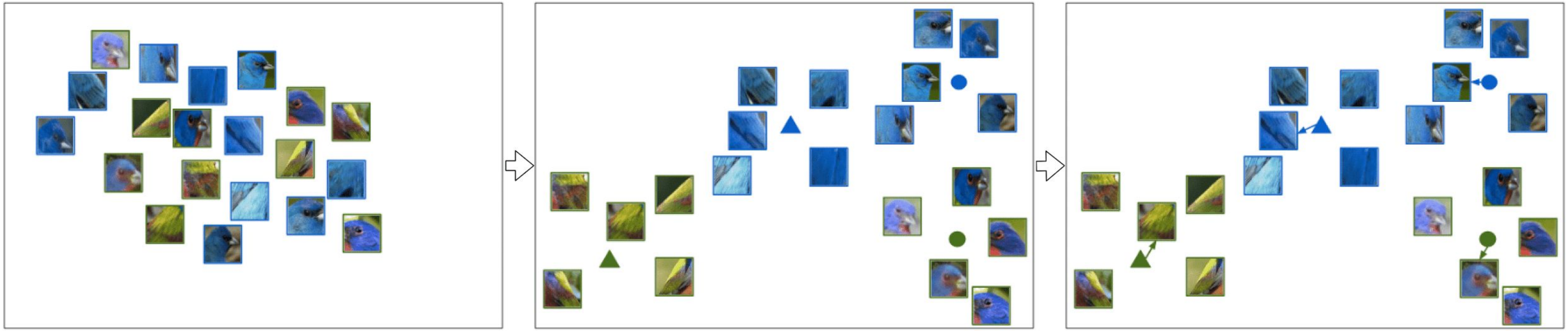
Sep: every latent patch of a training image stays away from prototypes of other classes

The training algorithm (2/3)

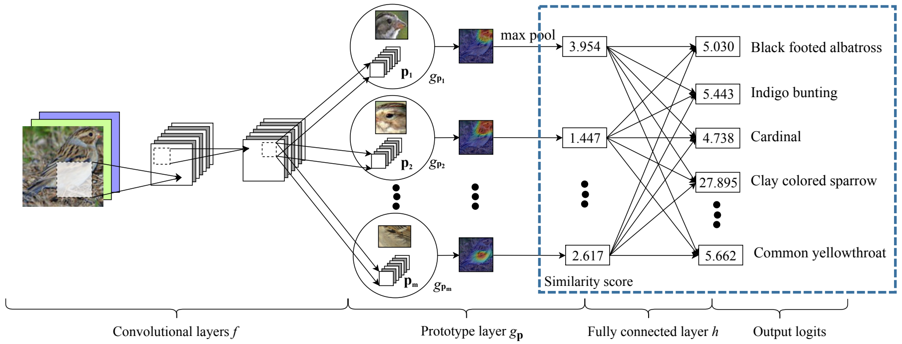
Stage 2: Projection of prototypes

$$\mathbf{p}_j \leftarrow \arg\min_{\mathbf{z}\in\mathcal{Z}_j} \|\mathbf{z}-\mathbf{p}_j\|_2, \text{where } \mathcal{Z}_j = \{\tilde{\mathbf{z}} : \tilde{\mathbf{z}} \in \text{patches}(f(\mathbf{x}_i))\ \forall i \text{ s.t. } y_i = k\}.$$

Each prototype projected onto the nearest latent training patch of the same class



The training algorithm (3/3)



Stage 3: Optimization of last layer

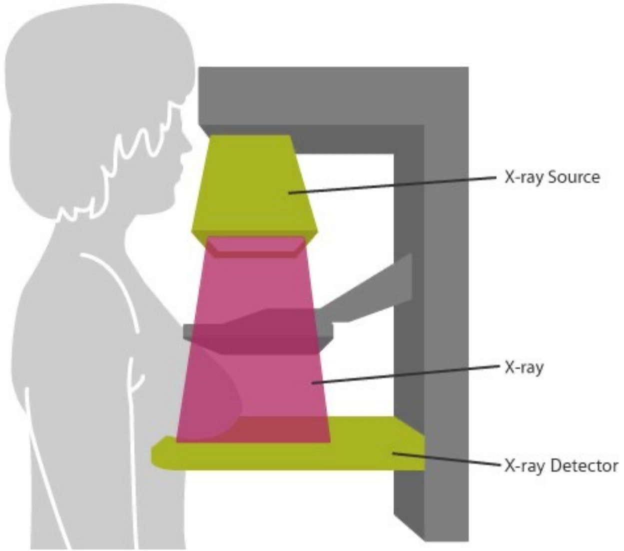
$$\min_{w_h} \frac{1}{n}\sum_{i=1}^{n} \mathrm{CrsEnt}(h \circ g_{\mathbf{p}} \circ f(\mathbf{x_i}), \mathbf{y_i}) + \lambda \sum_{k=1}^{K} \sum_{j:\mathbf{p}_j \notin \mathbf{P}_k} |w_h^{(k,j)}|.$$

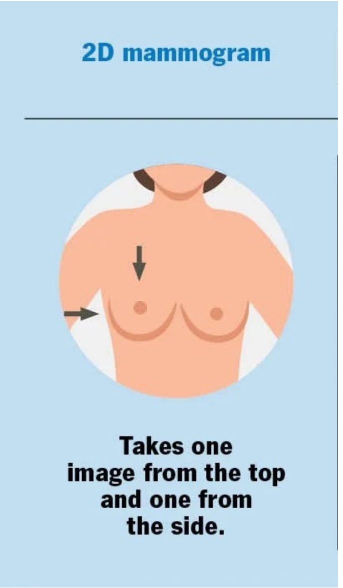
Adjust the last layer connection $w_h^{(k,j)}$ (k is the class index, j is the prototype index), so that for prototype p_j not in class k, $w_h^{(k,j)} \approx 0$

Less dependence on **negative** reasoning: "This does **not** look like that"

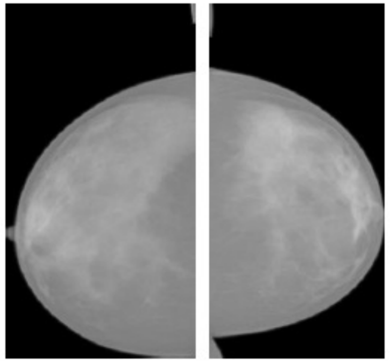
Mammography

- Low-energy X-ray acquisitions
- Two views – CC & MLO
- Breast tissue characterization



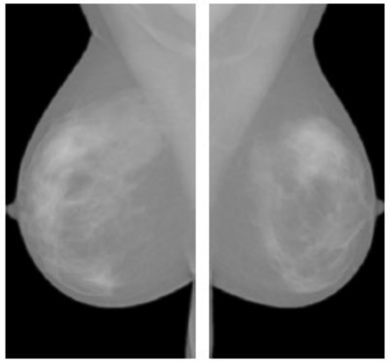


Cranio-Caudal (CC)



R L

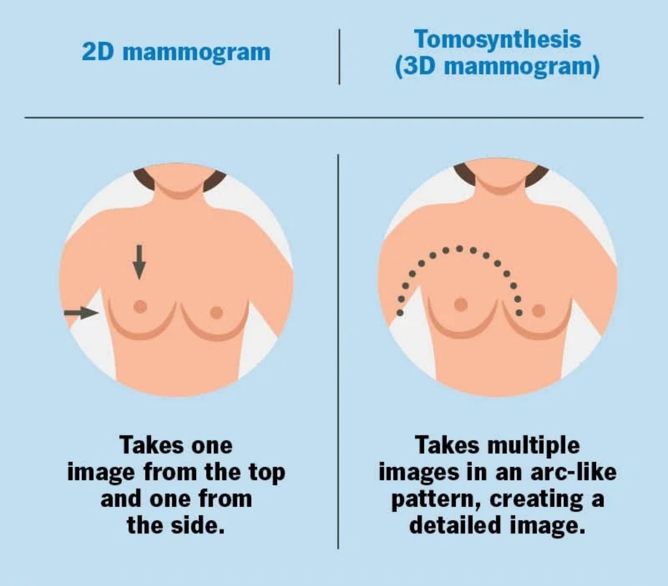
Medio-Lateral Oblique (MLO)



R L

Digital Breast Tomosynthesis

- Reduced tissue superposition
- More accurate cancer detection
- Particularly beneficial for dense breast tissue



https://my.clevelandclinic.org/health/diagnostics/15939-digital-breast-tomosynthesis-and-breast-cancer-screening

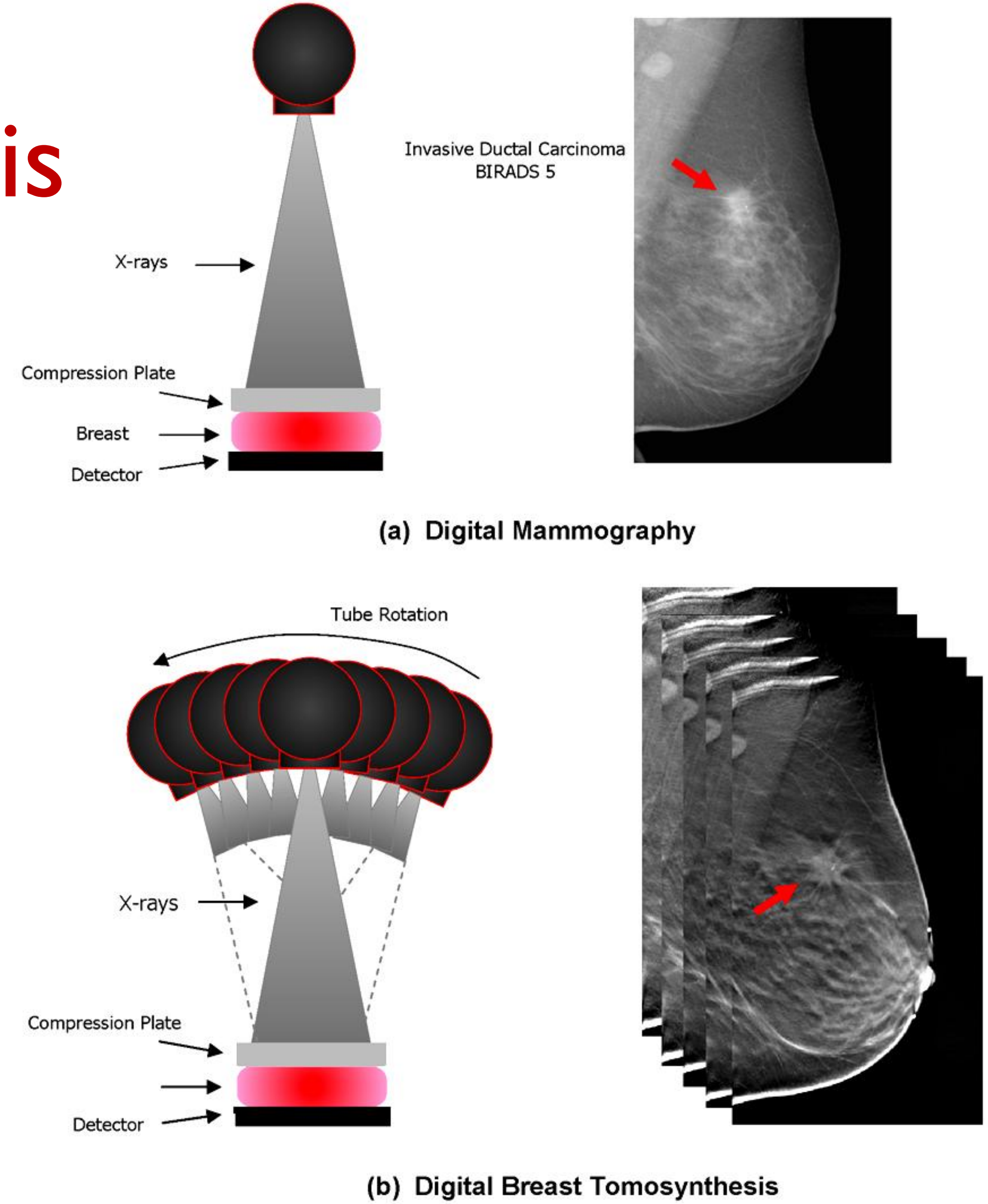


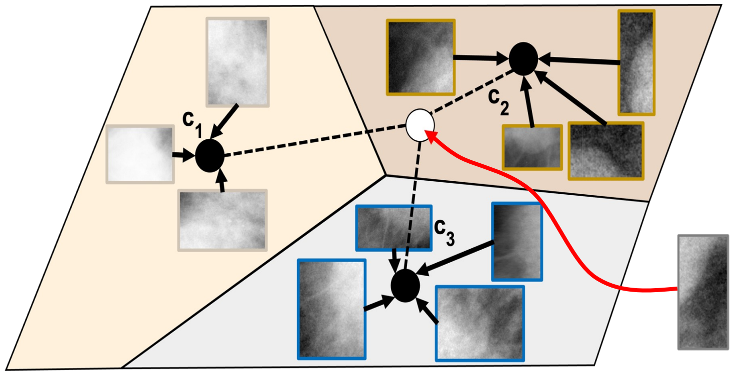
Image from: Kontos, D., Bakic, P. R., & Maidment, A. D. (2008, March). Texture in digital breast tomosynthesis: a comparison between mammographic and tomographic characterization of parenchymal properties. In *Medical Imaging 2008: Computer-Aided Diagnosis* (Vol. 6915, pp. 95-105). SPIE.

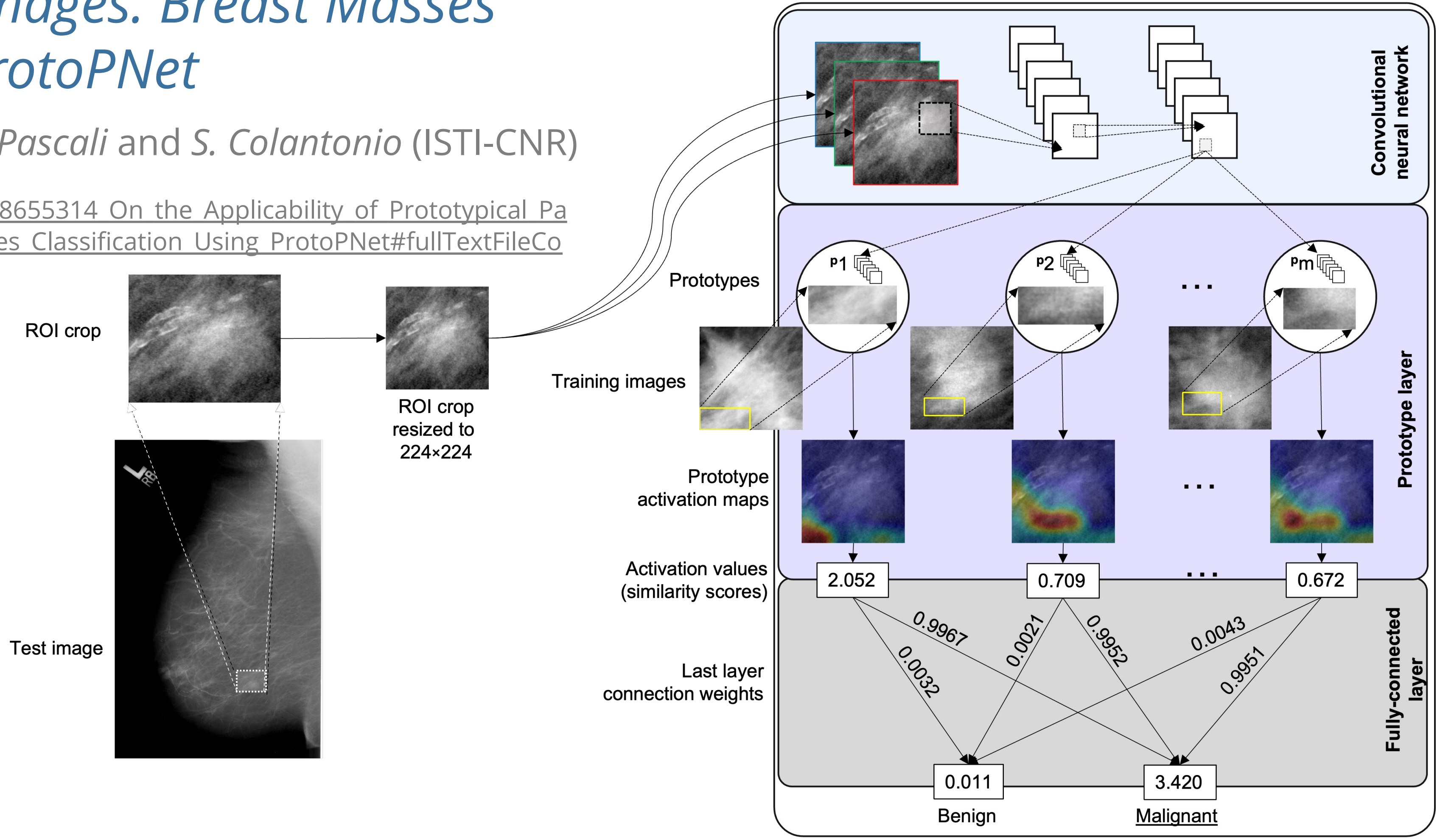
Our previous work

On the Applicability of Prototypical Part Learning in Medical Images: Breast Masses Classification Using ProtoPNet

with *G. Carloni, C. Iacconi, M. A. Pascali* and *S. Colantonio* (ISTI-CNR)

https://www.researchgate.net/publication/368655314_On_the_Applicability_of_Prototypical_Part_Learning_in_Medical_Images_Breast_Masses_Classification_Using_ProtoPNet#fullTextFileContent



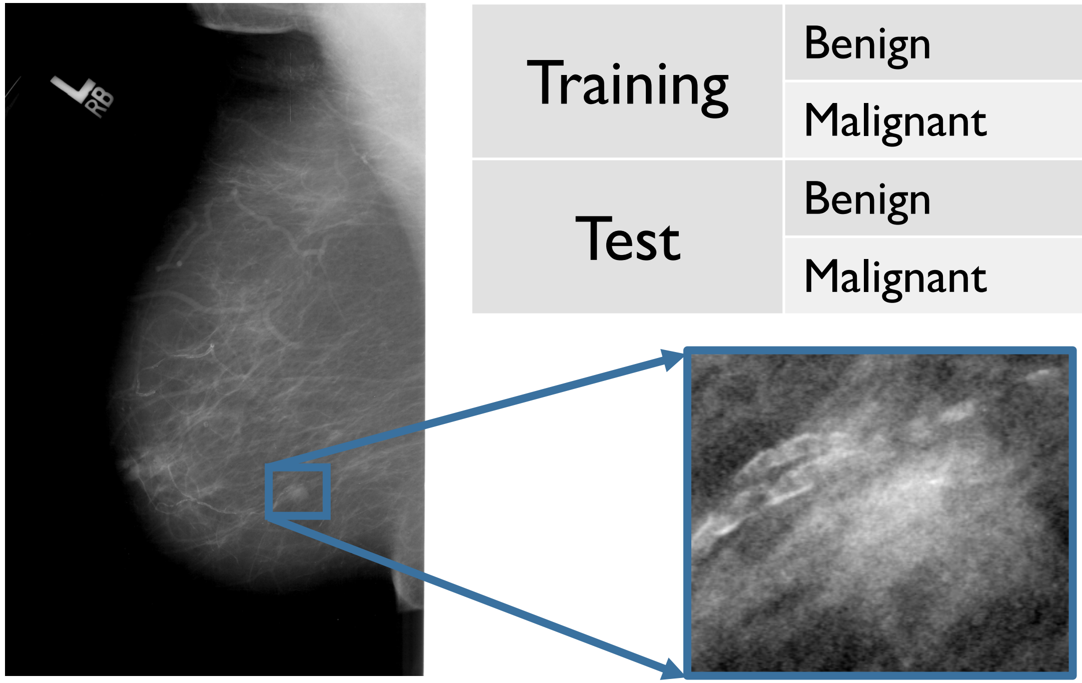


CBIS-DDSM Dataset

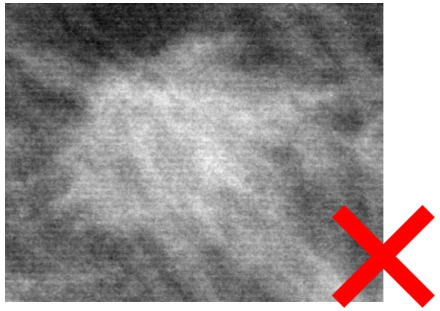
Dataset selection

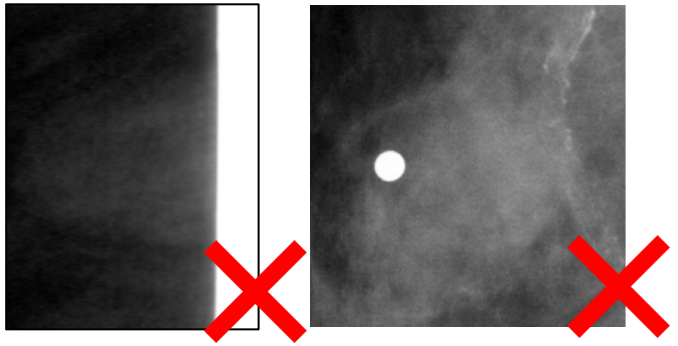
CBIS-DDSM (Curated Breast Imaging Subset of DDSM)

Cleaning and Balancing



Training	Benign	577	→	**528**
	Malignant	637	→	**528**
Test	Benign	194	→	**131**
	Malignant	147	→	**131**



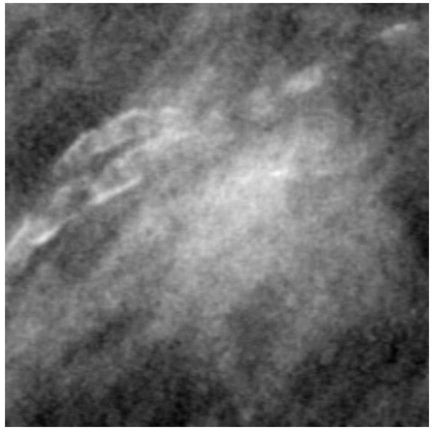


https://wiki.cancerimagingarchive.net/pages/viewpage.action?pageId=22516629

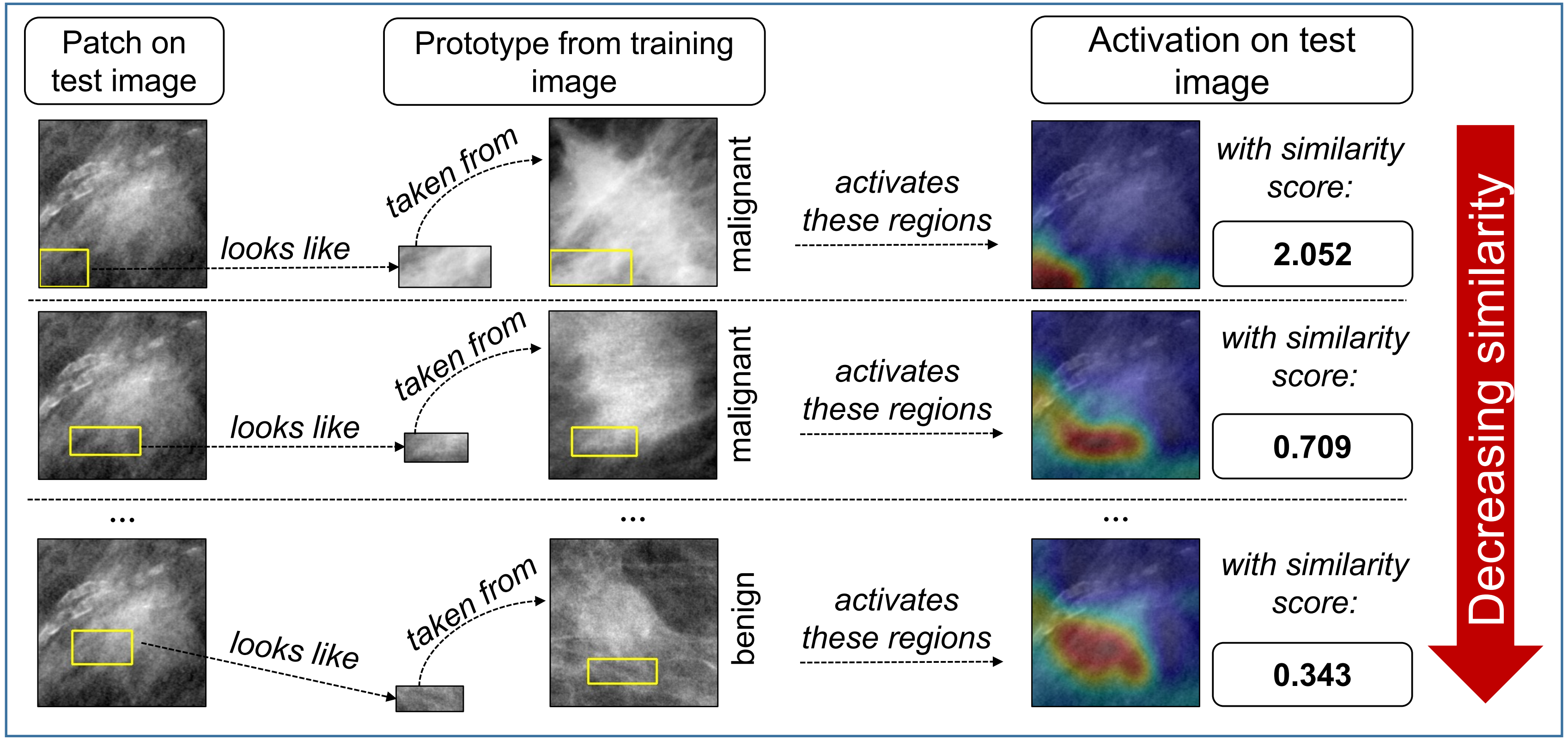
Differences from original ProtoPNet

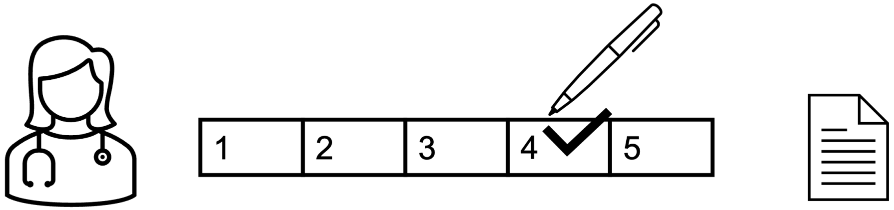
- Dataset:
  - From natural images to **medical images** → generation of **3-channel images** from single-channel medical images
- Training framework:
  - Presence of **hold-out test set**: assess the final performance, after training in **Cross-Validation**
  - **Fixed LR** value and **Early-stopping** during training process
- Architectural changes:
  - 2D **Dropout** and a 2D **Batch-norm** layer after each add-on convolutional layer
  - **Number of classes**: 2
- Clinical **feedback** on the quality of output **explanations**

Results of our previous work

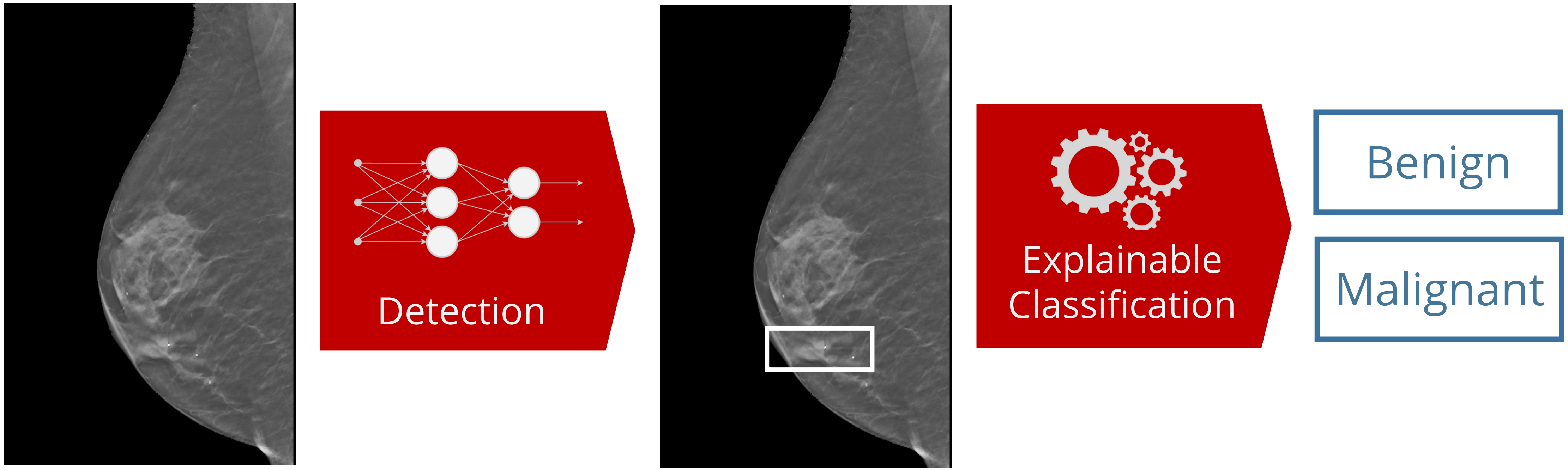


Test image: **malignant**
Predicted as: **malignant**

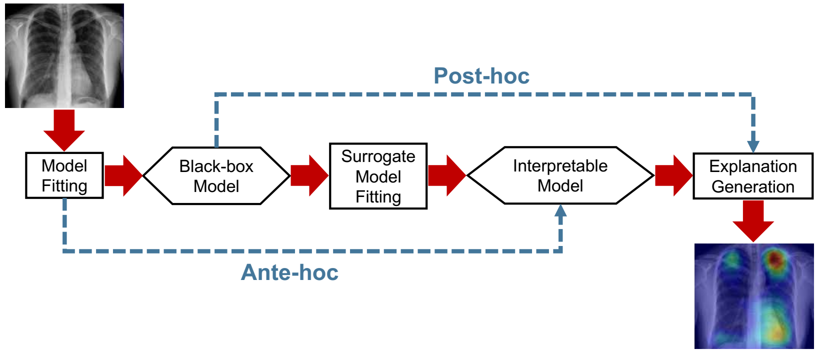




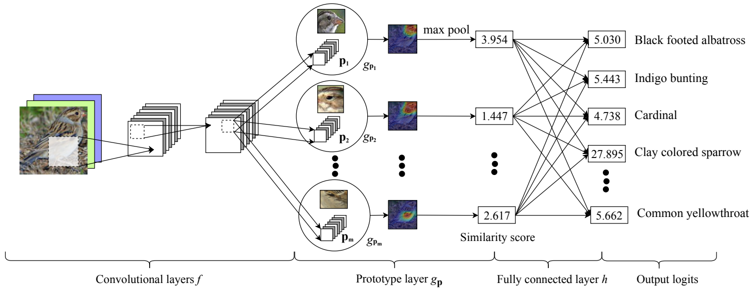
Our work on DBT images



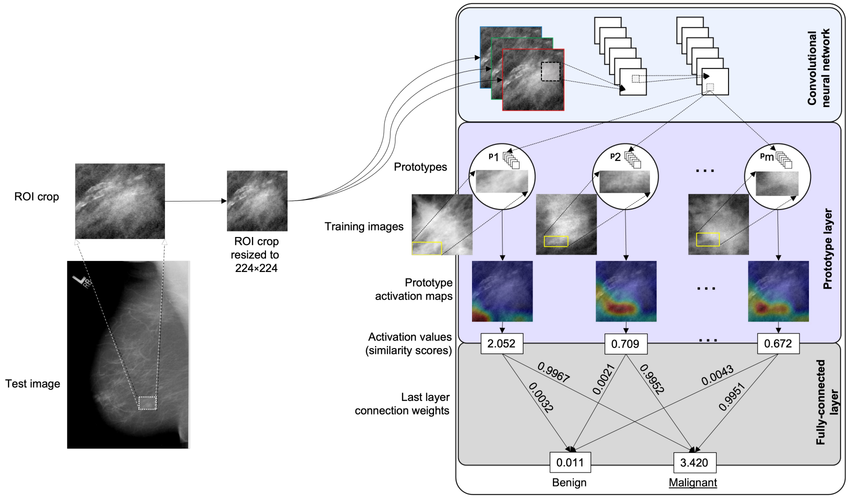
Conclusion



- Why explainability is important



- Explainability: post-hoc vs ante-hoc methods
- Case-based reasoning and ProtoPNet architecture
- ProtoPNet in medical imaging:
  - Mammography
  - Digital Breast Tomosynthesis



Thank You

Any Questions?

Andrea Berti
andrea.berti@isti.cnr.it

Ante-hoc explainability methods: the ProtoPNet architecture and its application on DBT images

ML-INFN

17/04/2023

Andrea Berti

Istituto di Scienza e Tecnologie dell'Informazione "A. Faedo"
Consiglio Nazionale delle Ricerche