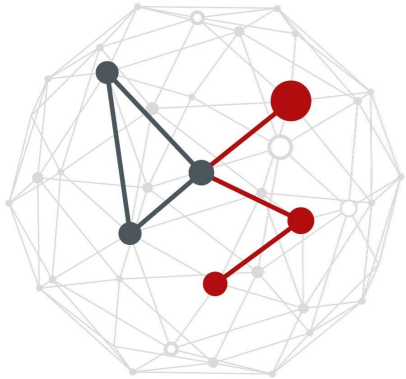
Master's degree ICT Internet Multimedia Engineering

VISUAL ATTENTION EXAMPLE



DIPARTIMENTO DI INGEGNERIA DELL'INFORMAZIONE

DIPARTIMENTO MATEMATICA

1222 · 2022 800 ANNI

UNIVERSITÀ DEGLI STUDI DI PADOVA

Visual attention

- Thorax disease classification on chest X-ray images
- General methods use global images of the thorax
- This has limitations
  - Generally the disease is evident within localized smaller areas
  - The use of global images may affect the classification performance
    - Excessive & irrelevant noisy regions
    - Poor alignment of diseased portions, that may be located along the borders
- Solution from [Guan2018] Attention Guided CNN (AG-CNN)

[Guan2018] Qingji Guan, Yaping Huang, Zhun Zhong, Zhedong Zheng, Liang Zheng, Yi Yang, “Diagnose like a radiologist: Attention guided convolutional neural network for thorax disease classification,” arXiv preprint arXiv:1801.09927.

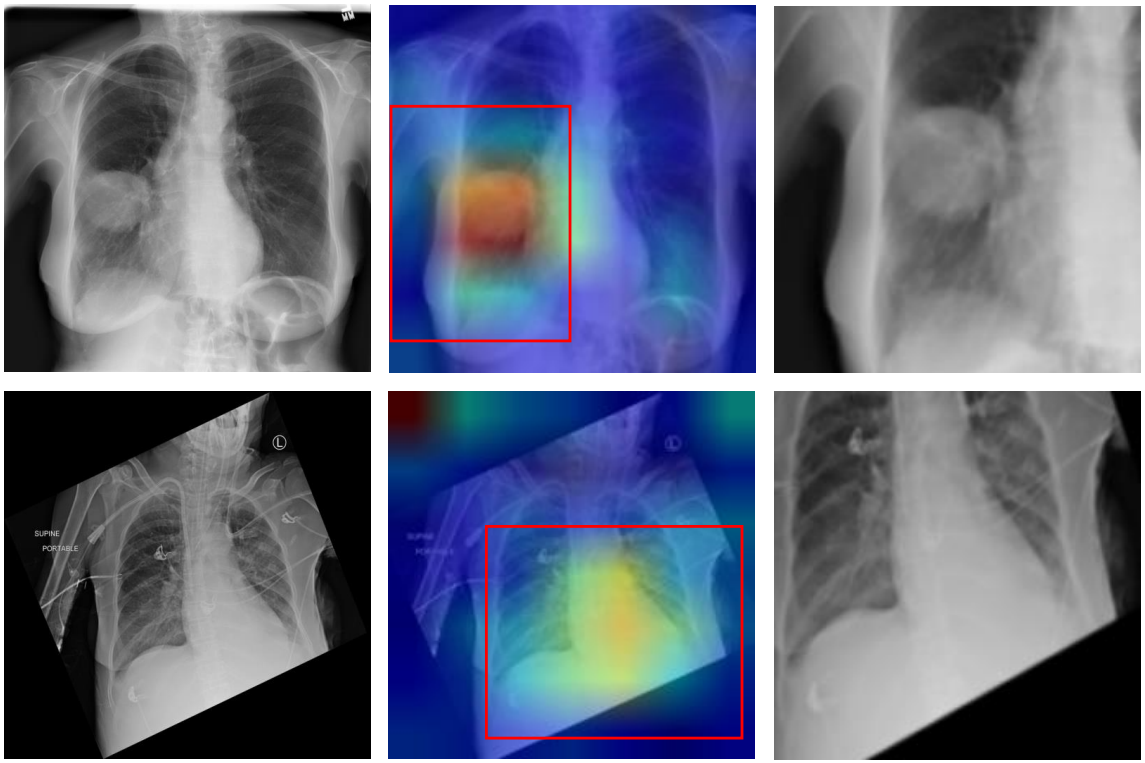
- Code on Github: https://github.com/Ien001/AG-CNN

Visual insight

(a) Original global X-ray image

(b) Heatmap is extracted from original global image

(b) The relevant sub-regions are cropped into local images



(a) original global image (b) heatmap (c) cropped local image

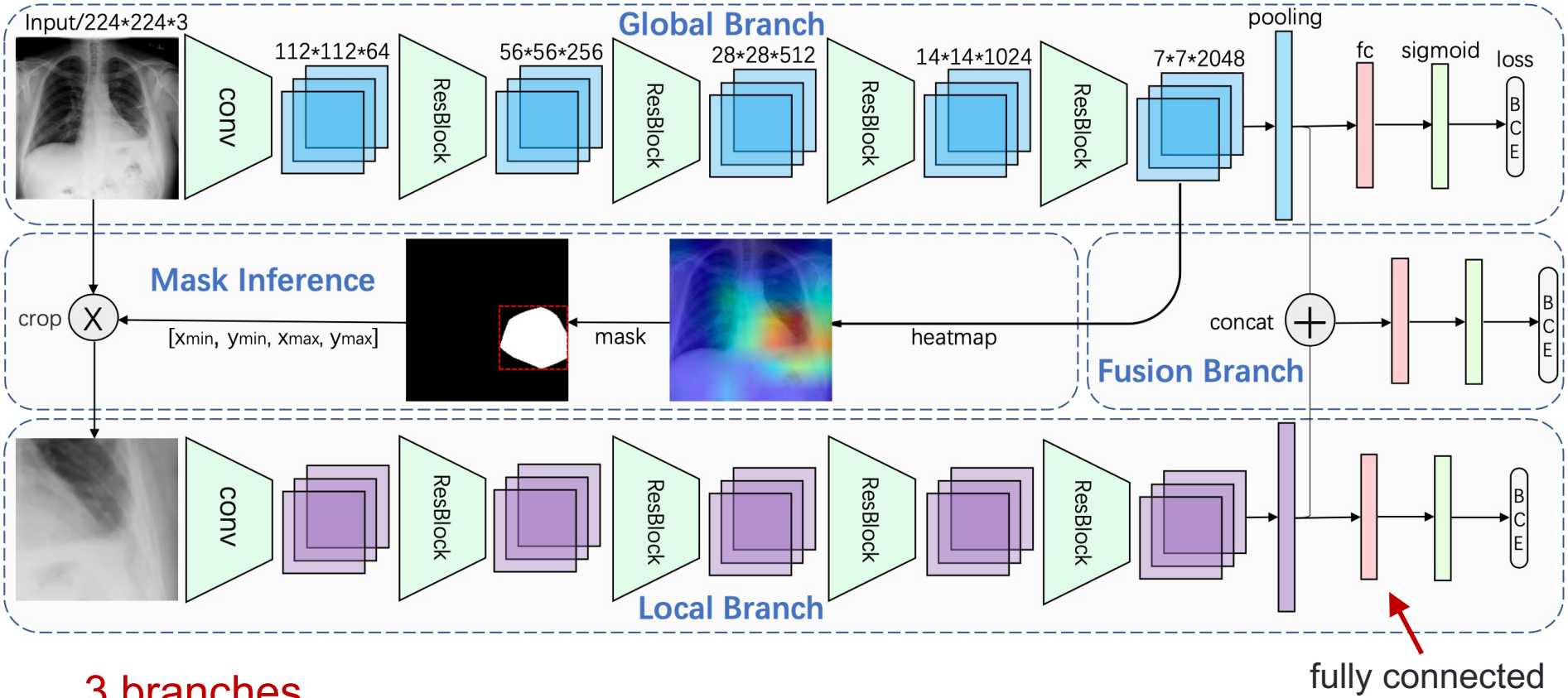
Dataset & labels

- Labeled dataset of X-ray chest images
  - 15 disease types (classes)
- Each image is labeled with a 15-dimensional vector

$$\boldsymbol{\ell} = [\ell_1\, \ell_2\, \cdots\, \ell_{15}]^\top,\ \ell_i \in \{0, 1\}$$

  - each entry: 0 if corresponding pathology is absent and 1 otherwise
- Pathologies (15 classes)
  - atelectasis, cardiomegaly, effusion, infiltration, mass, nodule, pneumonia, pneumothorax, consolidation, edema, emphysema, fibrosis, pleural thickening, hernia
- ChestX-ray14 dataset
  - https://www.rsna.org/en/education/ai-resources-and-training/ai-image-challenge/RSNA-Pneumonia-Detection-Challenge-2018

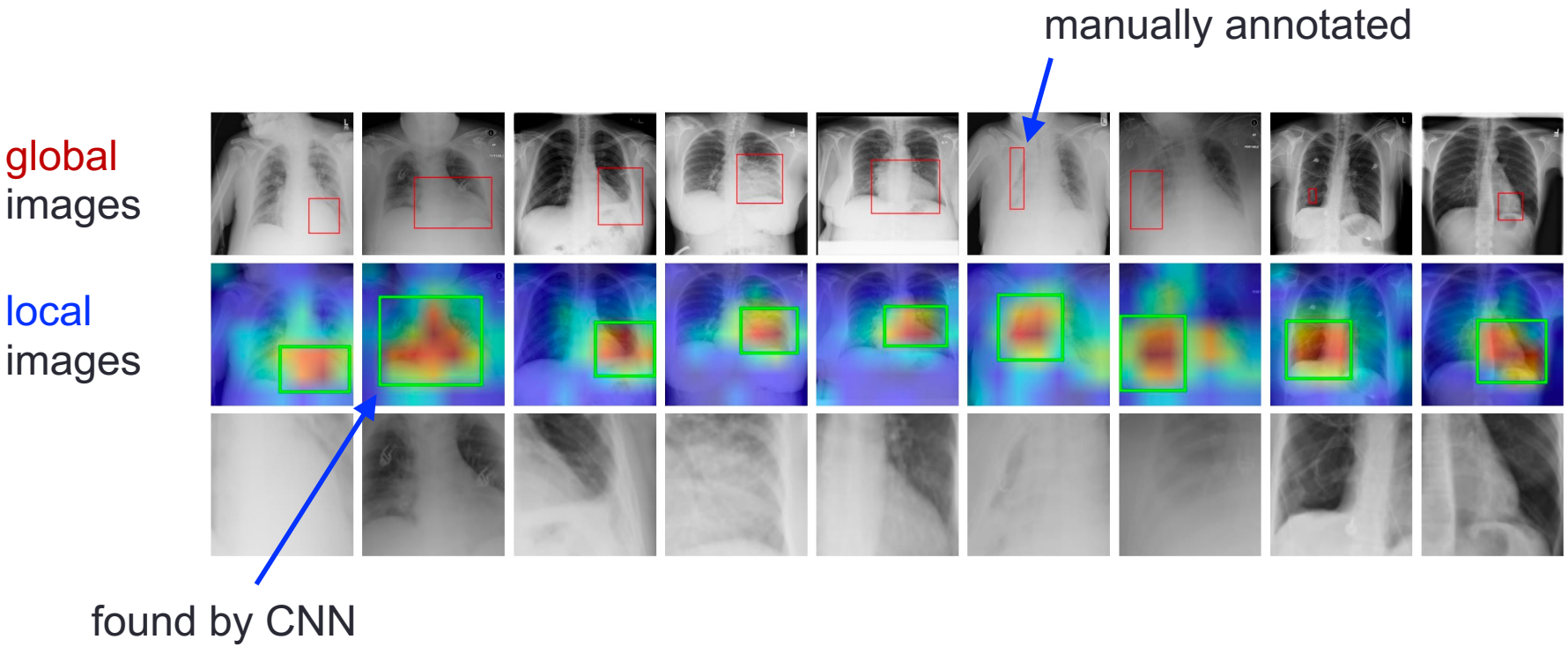
Neural network: based on ResNet



3 branches

- Global branch: learns to attend to a certain area (problem area)
- Local branch: performs classification using the cropped local area
- Fusion branch: uses *global* and *local* features to output the final class

Results of “global branch CNN”



- Global images: are manually annotated (see the red bounding boxes)
- Local images: are found by the “global branch CNN”

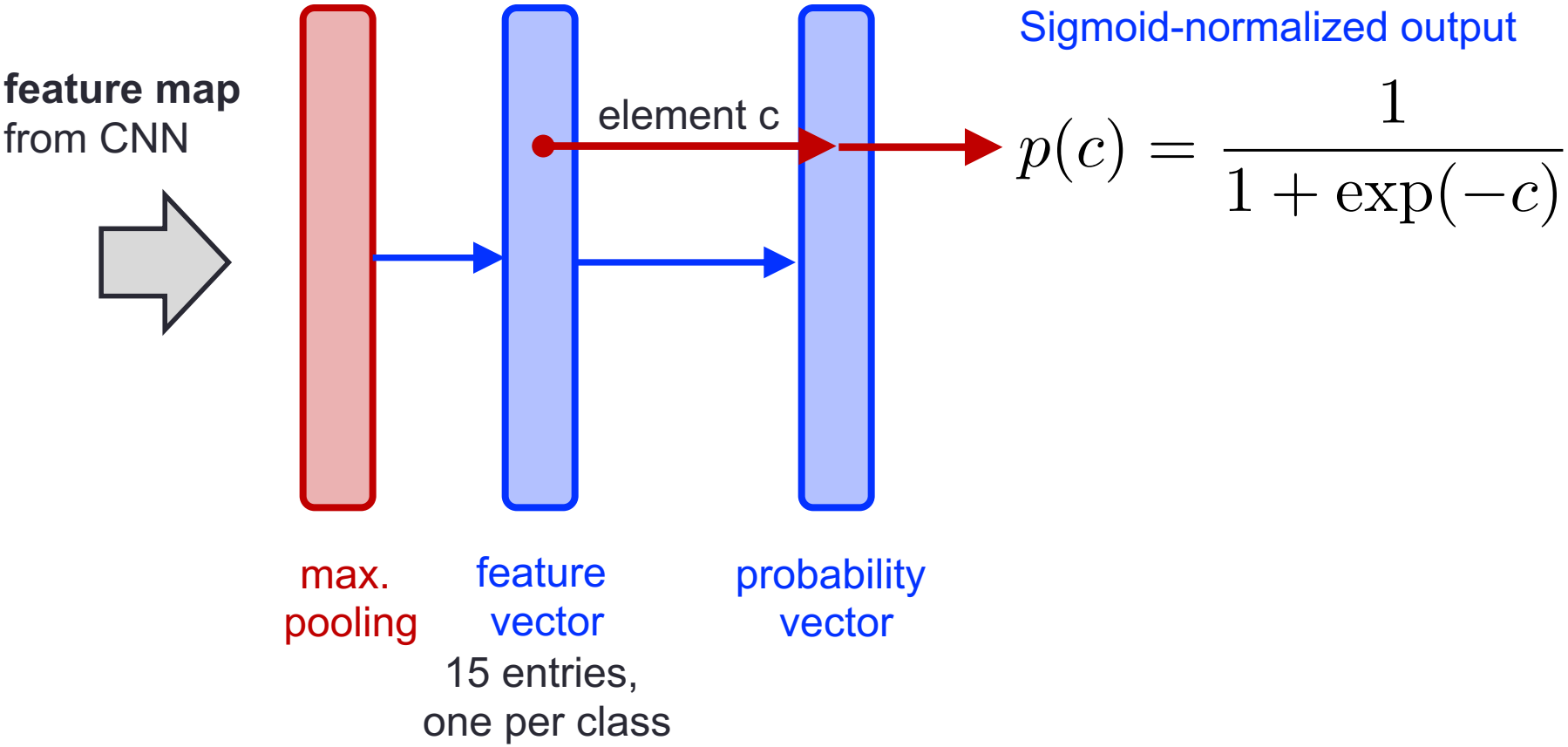
Important: the annotated regions (red boxes on top images) are neither used for learning, nor for testing purposes

Training procedure – step 1

- Global Branch CNN
  - Based on ResNet design
  - Processes the entire image and determines the region of interest (ROI) → to be cropped
  - The network weights are learned in a supervised manner using *backpropagation* and the *disease labels*
- When learning is complete
  - The feature map right before the final pooling layer is extracted and used to crop the ROI (according to the intensities of the pixels in such feature map)

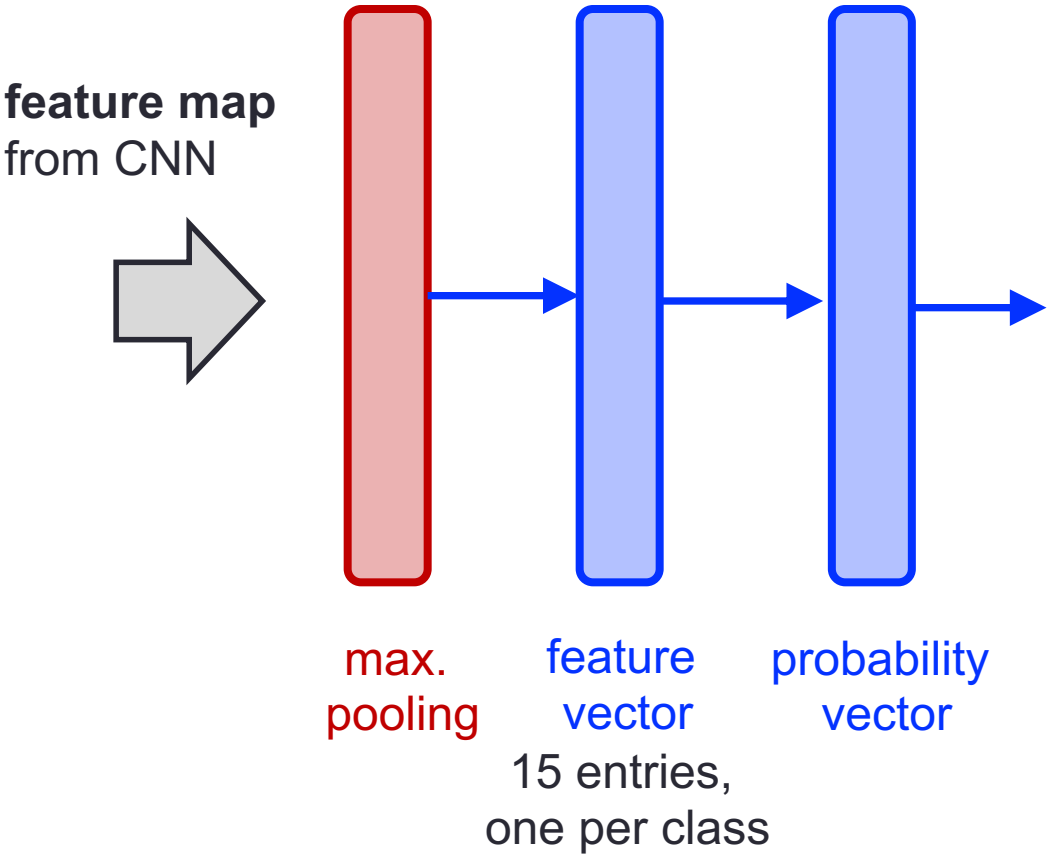
Output of global branch

- Pooling → feature vector (15 entries = num. of classes)



Output of global branch

- Pooling → feature vector (15 entries = num. of classes)



Binary Cross-Entropy (BCE) loss

$$L = \sum_{c=1}^{15} \Big[\ell_c \log(p(c)) + (1-\ell_c) \log(1-p(c)) \Big]$$

Training procedure – step 2

- Local Branch CNN
  - Is also based on a ResNet design
  - Same max. pooling, feature vector and BCE loss as in step 1
  - It takes as input the cropped ROI
  - Its weights are trained in a supervised manner (backprop, after step 1 and **NOT** jointly with it) using again the labels from the dataset
  - Uses entirely independent weights wrt the Global Branch CNN
  - Trained **after training** the Global Branch CNN

Training procedure – step 3

- Fusion branch
  - Is a fully connected layer
  - Same max. pooling, feature vector and BCE loss as in steps 1 & 2
  - Again, the weights are independent of those of the other branches
  - Uses the features extracted by the global and the local branches
    - Fusing features to obtain better classification results
    - Funsion layer: concatenates the features (**pooling output**) from the *global* and *local* branches, then a fully connected layer is applied to this concatenated vector → obtaining the 15-dimensional output vector. This last vector is passed (element-wise) through a sigmoid nonlinearity to obtain the 15 output probabilities for the considered diseases
  - Trained **after training** Global and Local Branches