# Zoom-in-Net: Deep Mining Lesions for Diabetic Retinopathy Detection

Journal Club 26/06/17

**1230x1230**


By providing image-level labels can we 1) determine where the lesions are and 2) accurately label unseen test images

1230x1230


By providing image-level labels can we 1) determine where the lesions are and 2) accurately label unseen test images

Can we use local information to aid image-level classification

Classification

Localization

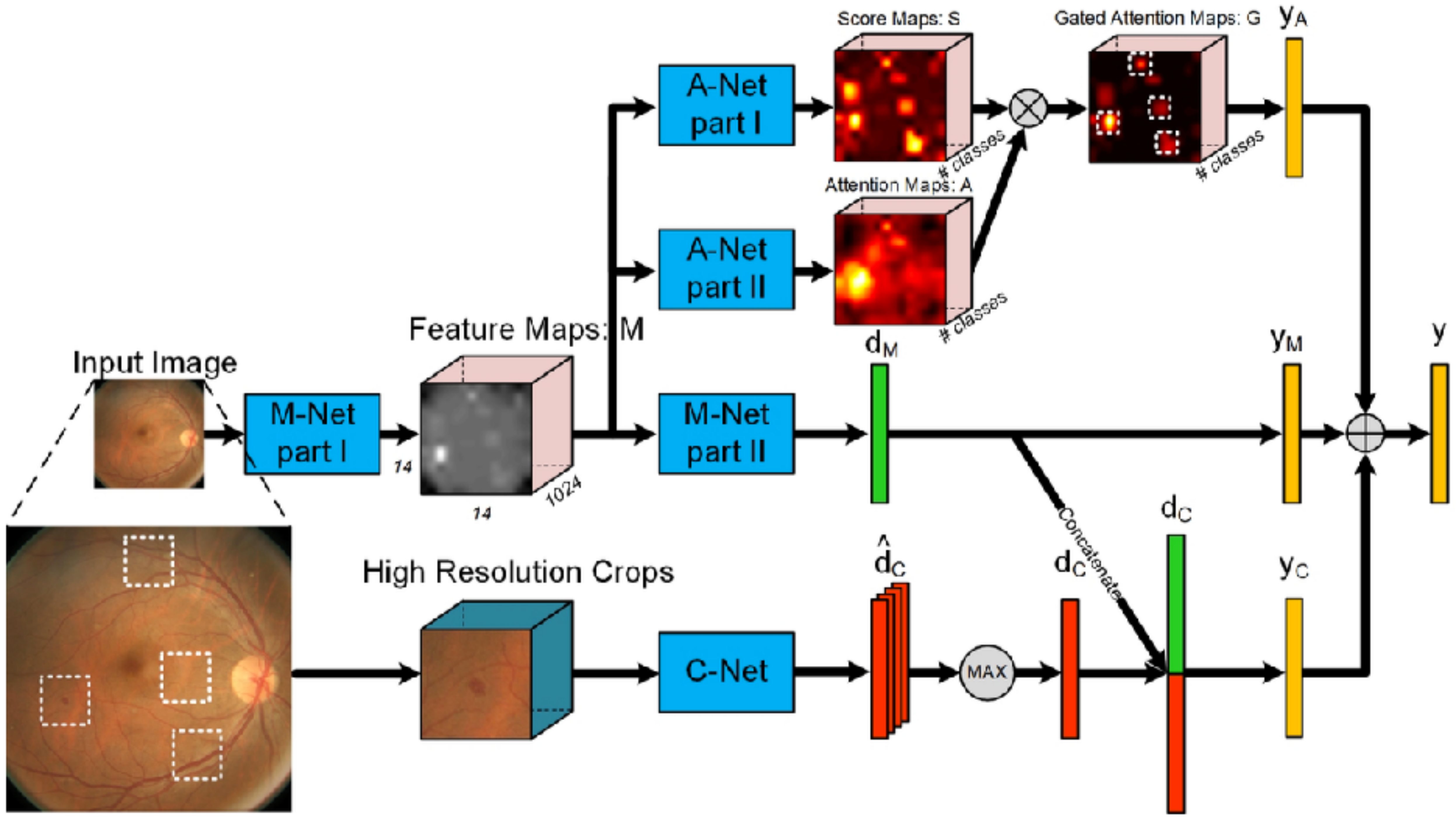



**Fig. 1.** An overview of Zoom-in-Net. It consists of three sub-networks. M-Net and C-Net classify the image and high resolution suspicious patches, respectively, while A-Net generates the gated attention maps for localizing suspicious regions and mining lesions.

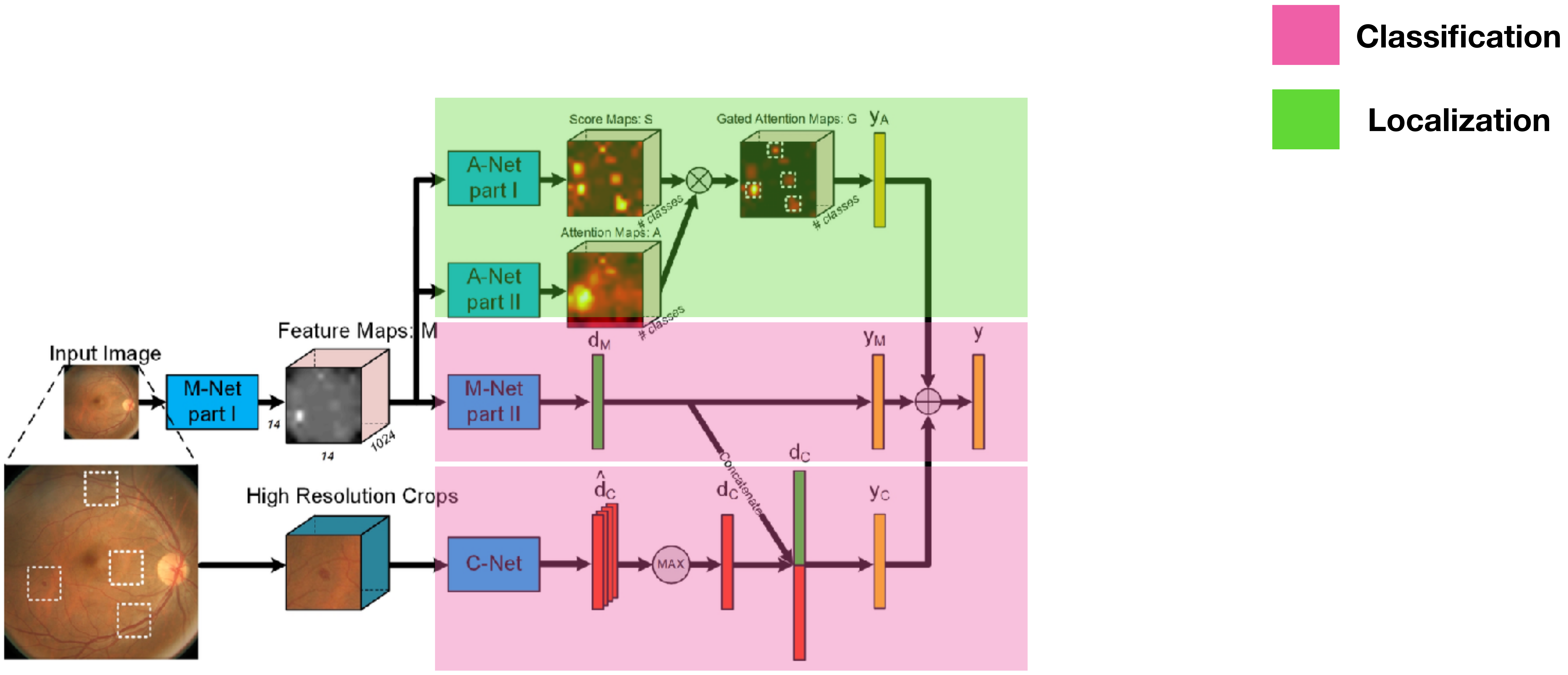



**Fig. 1.** An overview of Zoom-in-Net. It consists of three sub-networks. M-Net and C-Net classify the image and high resolution suspicious patches, respectively, while A-Net generates the gated attention maps for localizing suspicious regions and mining lesions.

Traditional CNN structure to extract features

- Main Network (M-Net)
- Attention Network (A-Net)
- Crop Network (C-Net)

Localize regions-of-interest with probabilistic scores

input

# M-Net

- Inception-**ResNet**

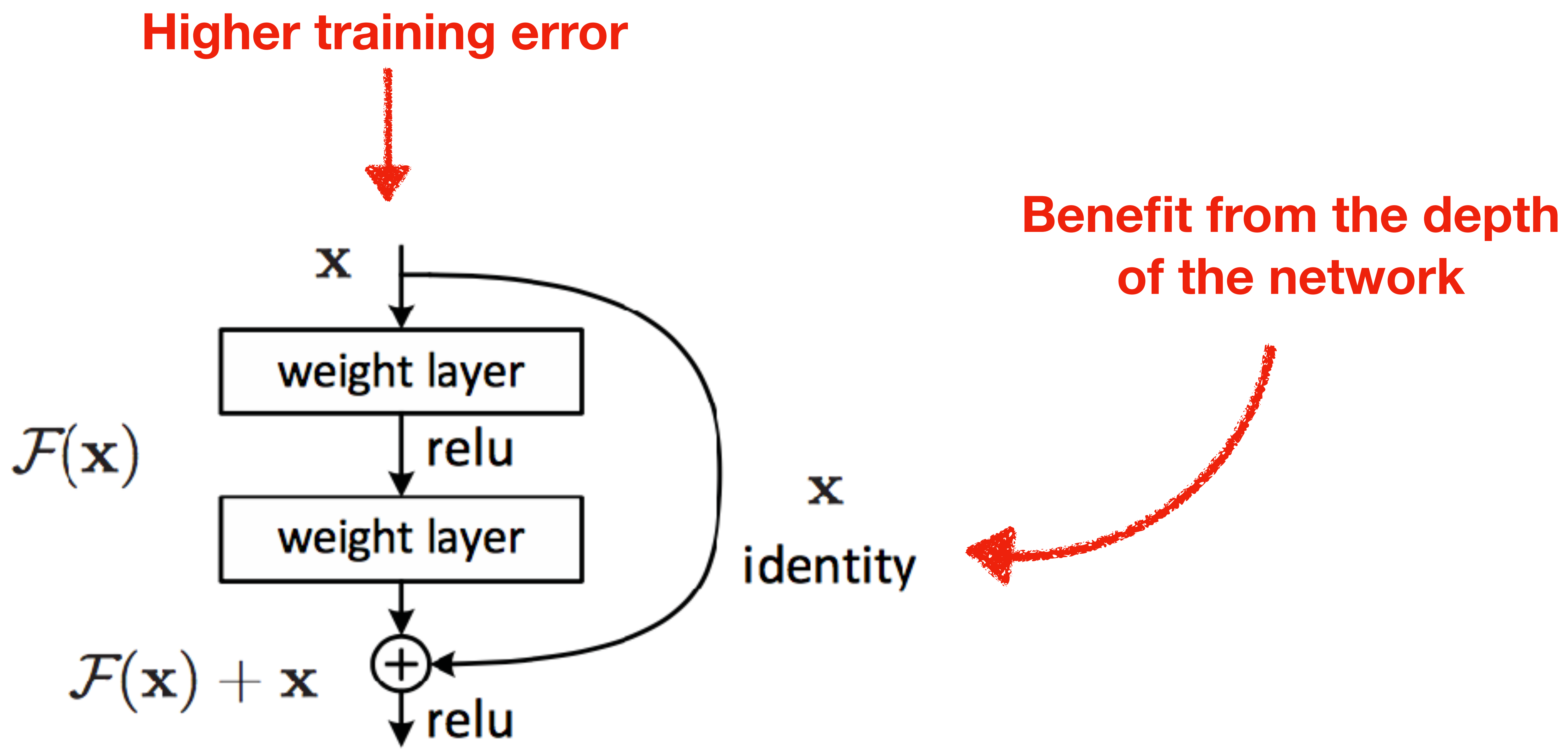



Figure 2. Residual learning: a building block.

# M-Net

- Inception-**ResNet**

Higher training error

Benefits from the depth of the network

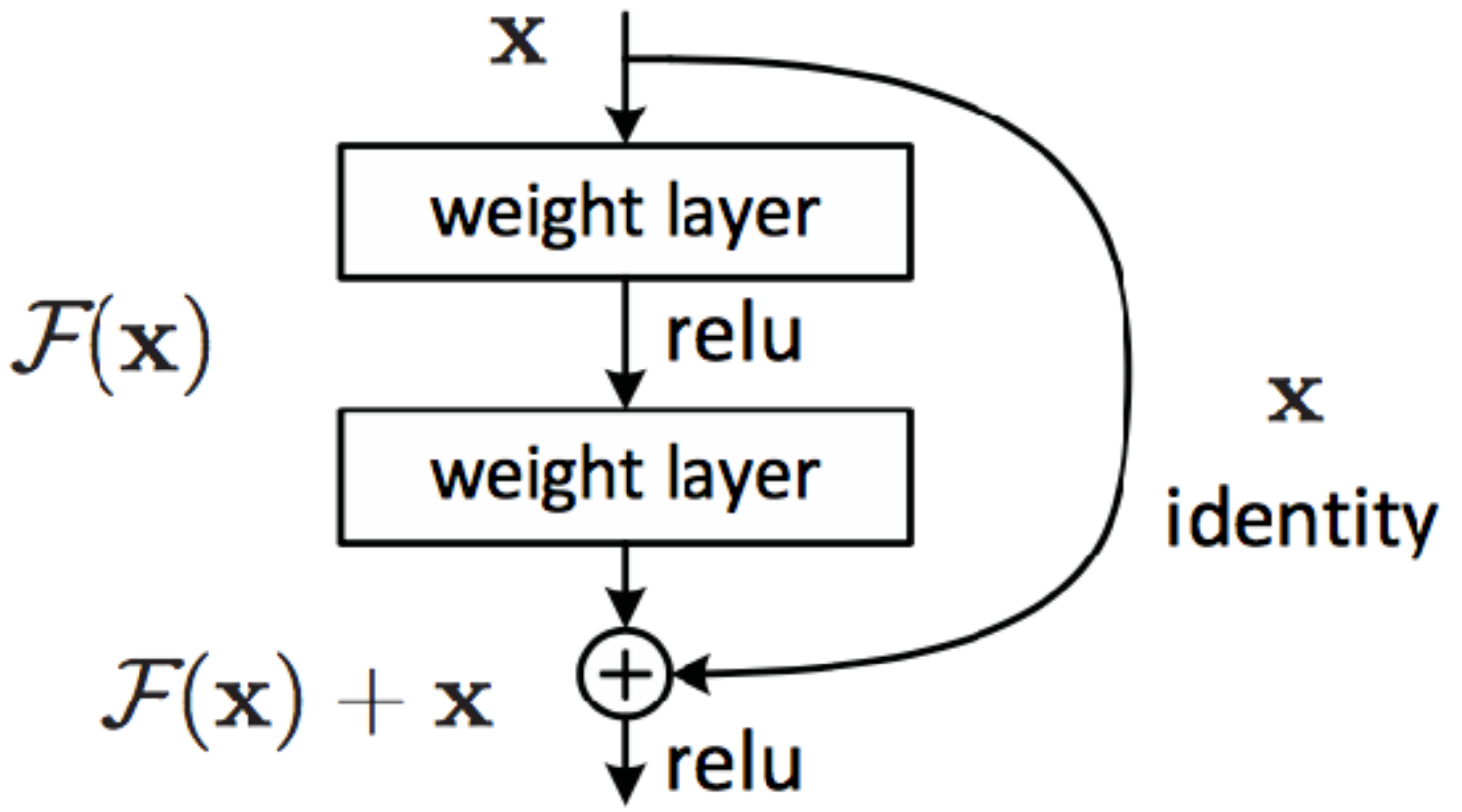



Figure 2. Residual learning: a building block.

The authors argue that residual connections are inherently necessary for training very deep convolutional models. Our findings do not seem to support this view, at least for image recognition

# M-Net

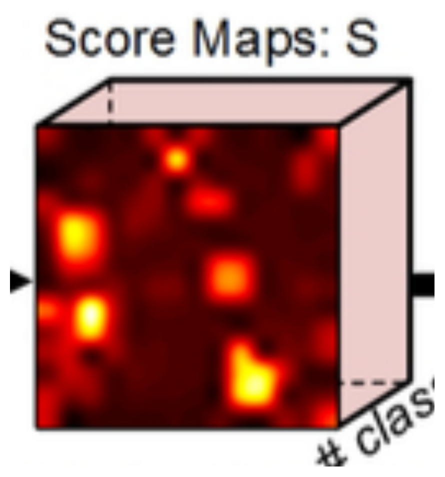
- Inception-**ResNet**
- They use the same network for two different reasons:
  1. First, to perform disease level classification (5 levels)
  2. Second, to pass extracted features to C-Net

# Zooming-In (A-Net)

**M-Net features as input…**

1x1 Conv

- Score maps



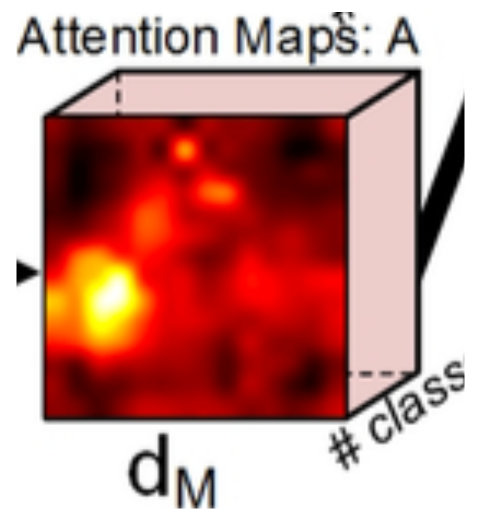

**Pixel-level predictions**

3x3 Conv

3x3 Conv

1x1 Conv

- Attention gate maps



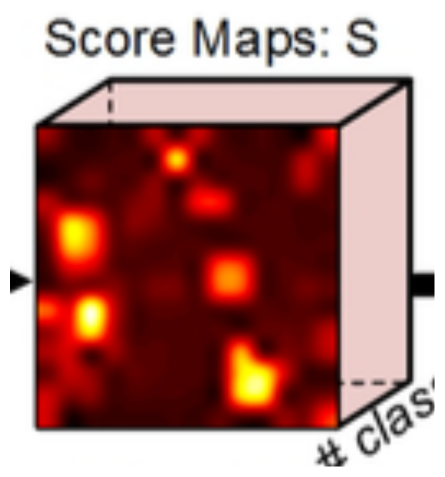
# Zooming-In (A-Net)

**M-Net features as input…**

1x1 Conv

**Pixel-level Prediction**

- Score maps




**Pixel-level predictions**

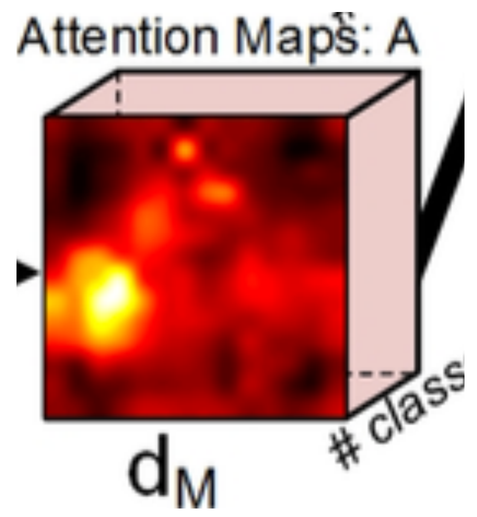
3x3 Conv

3x3 Conv

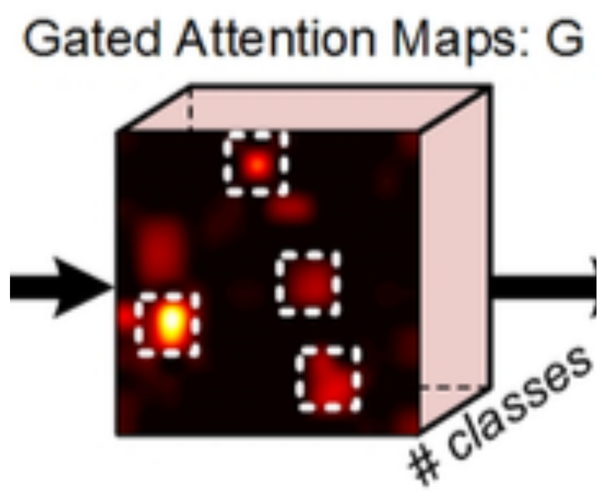
1x1 Conv

- Attention gate maps

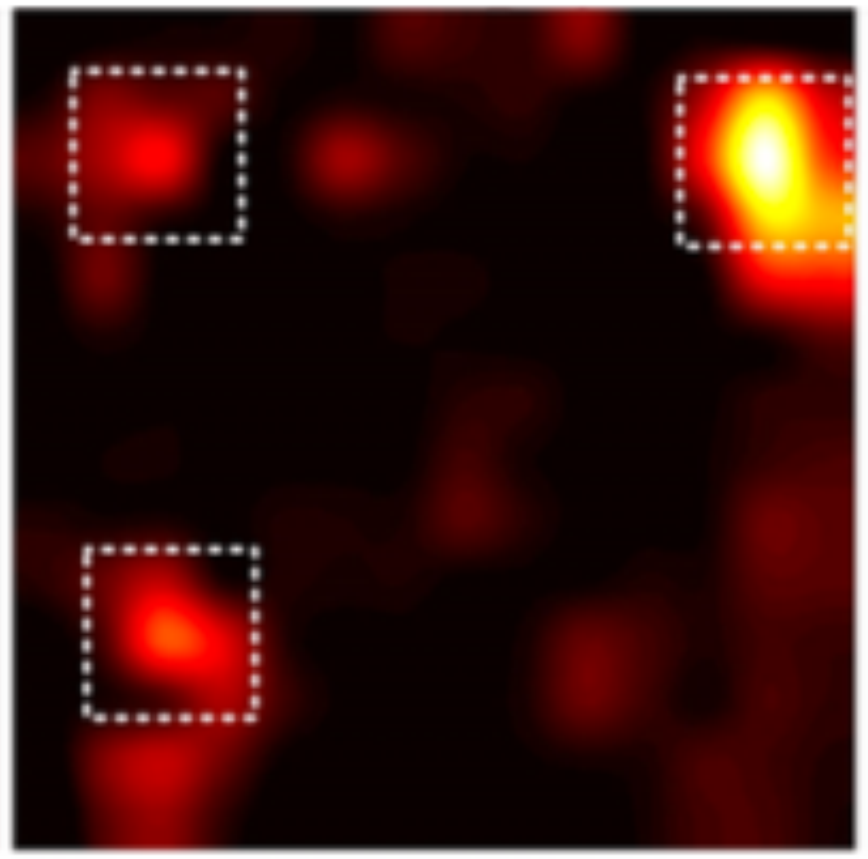
**Disease-level Prediction**



# Zooming-In (C-Net)

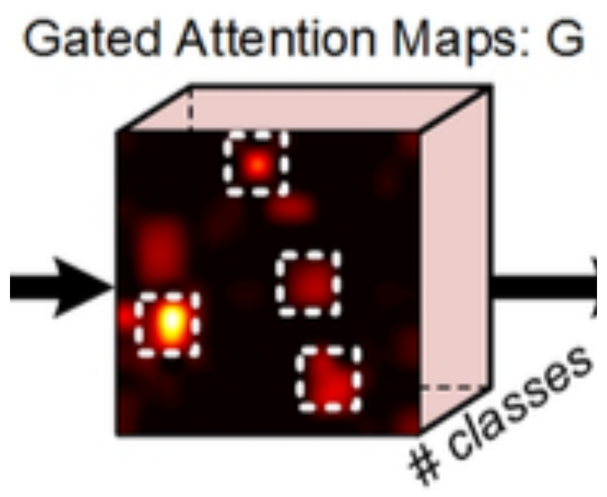



**Resize to original size of image**

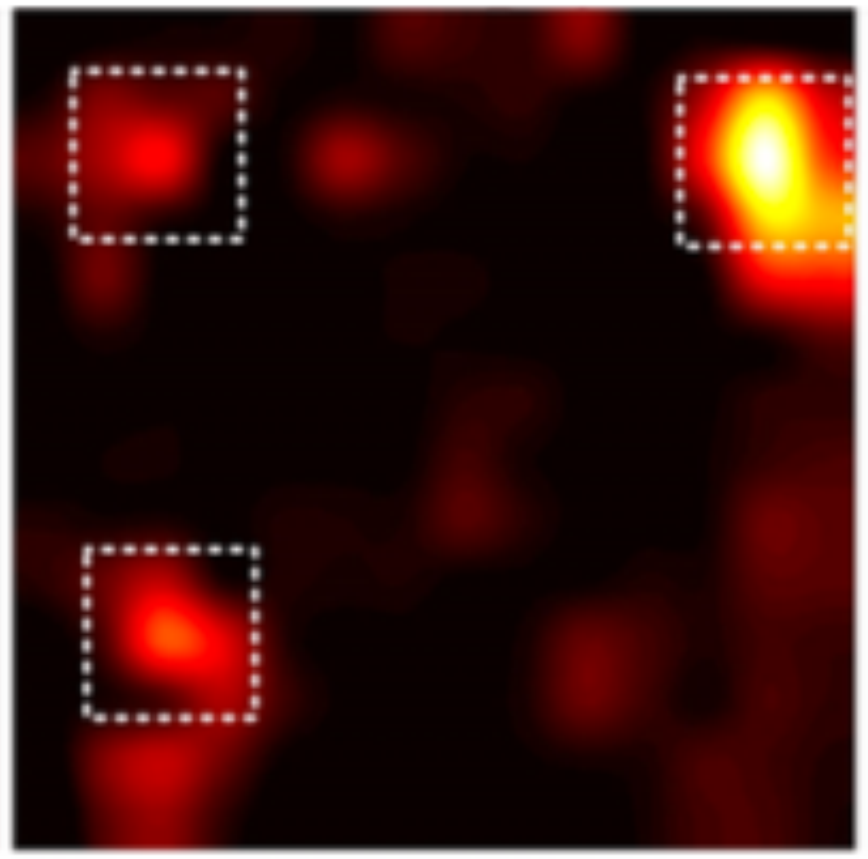


**Select top N responses which are then extracted for training C-Net**

# Zooming-In (C-Net)

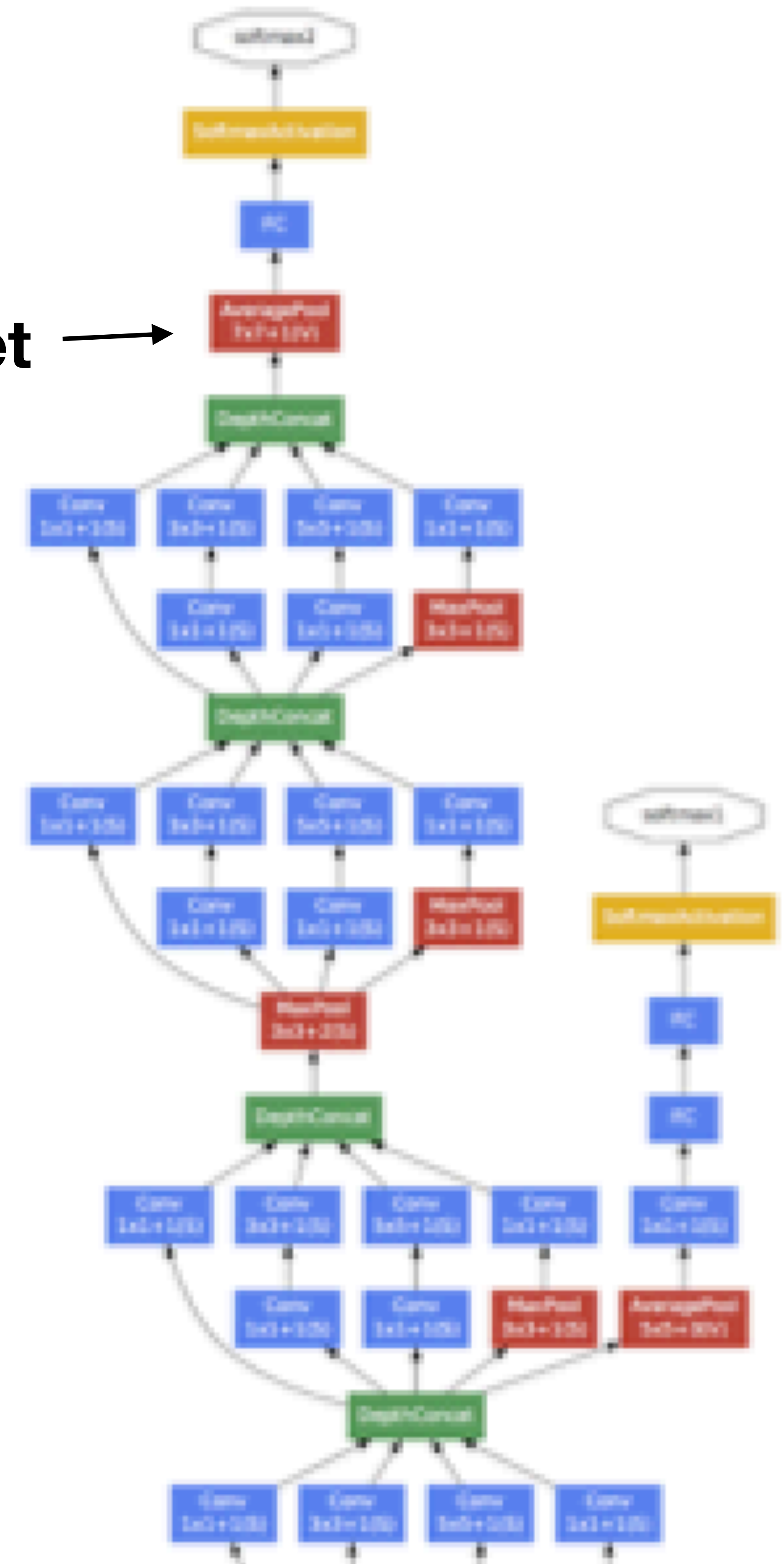



Resize to original size of image



Select top N responses which are then extracted for training C-Net

Add features from M-Net



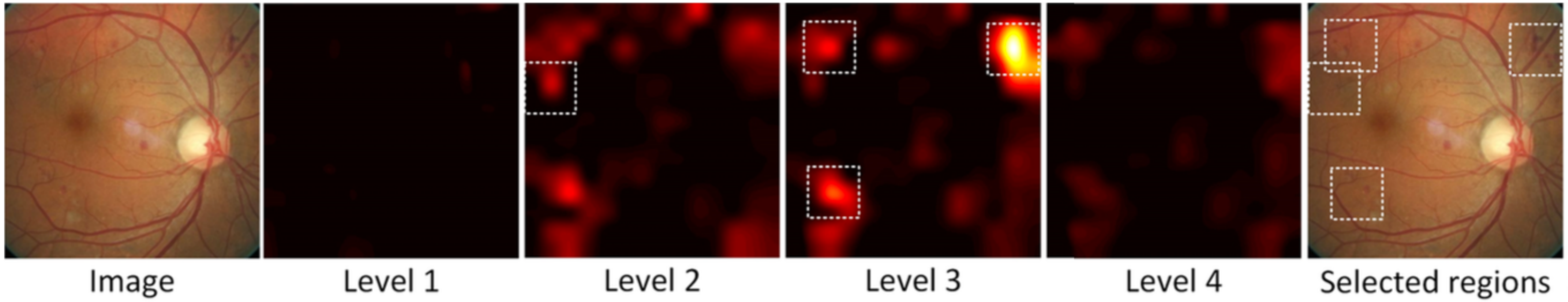
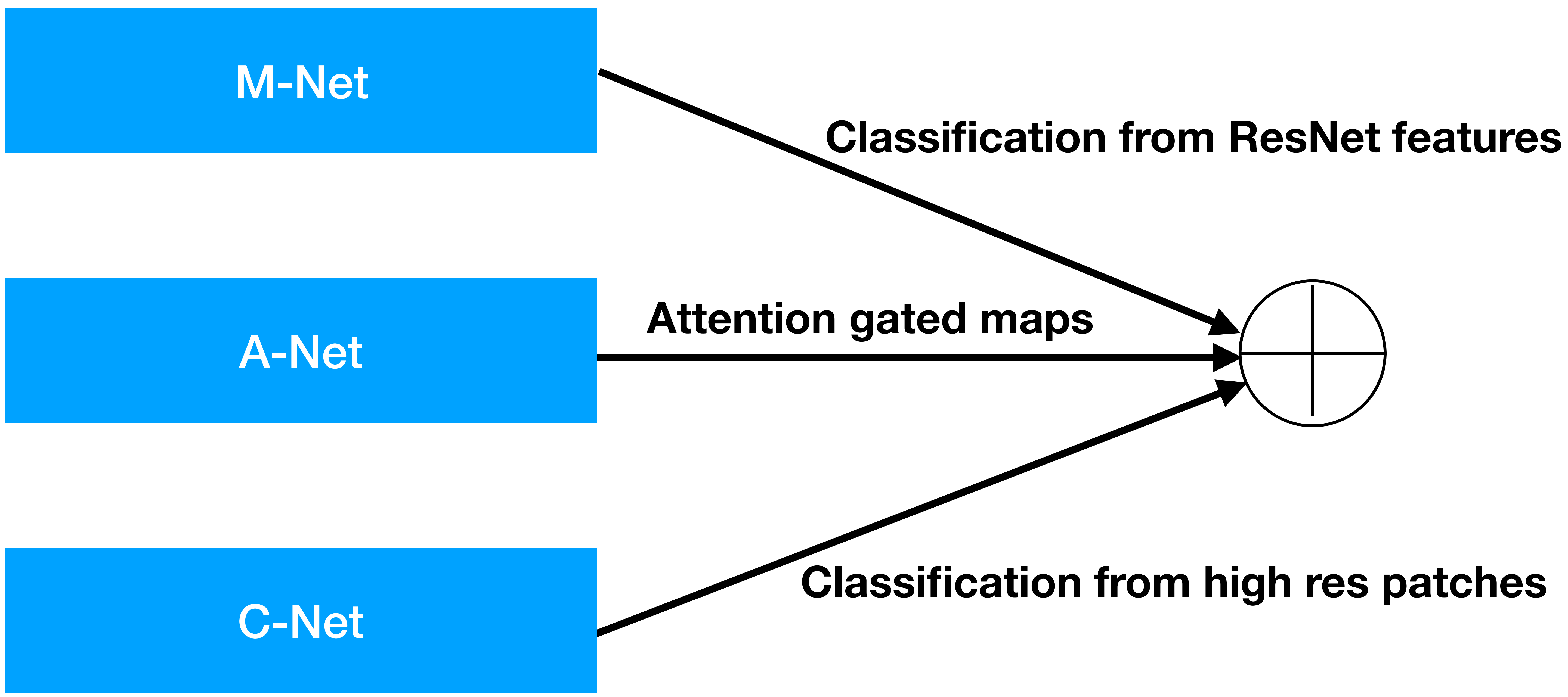



**Fig. 3.** From left to right: image, gated attention maps of level 1-4 and the selected regions of the image. The level 0 gated attention map has no information and is ignored.


M-Net
A-Net
C-Net
Classification from ResNet features
Attention gated maps
Classification from high res patches

# Evaluation

“Only four bounding boxes generated from the automatically learned attention maps are enough to cover 80% of the lesions labeled by an experienced opthalmologist”

- Dataset was augmented (via rotations)
- M-Net was pretrained with ImageNet
- Learning rate was decreased during training

| Algorithms | val set | test set |
|---|---|---|
| Min-pooling [1] | 0.86 | 0.849 |
| o_O | 0.854 | 0.844 |
| Reformed Gamblers | 0.851 | 0.839 |
| M-Net | 0.832 | 0.825 |
| M-Net+A-Net | 0.837 | 0.832 |
| Zoom-in-Net | 0.857 | 0.849 |
| Ensembles | **0.865** | **0.854** |

**Table 2.** Comparison to top-3 entries on Kaggle' challenge.

Task 1: Non referable (Grade 0/1), Referable (Grade 2/3)

Task 2: Normal vs Abnormal

SVM was trained on EyePACS and then tested on Messidor

If classified as level 0 then normal

| Method | AUC | Acc. |
|---|---|---|
| Lesion-based [12] | 0.760 | - |
| Fisher Vector [12] | 0.863 | - |
| VNXK/LGI [18] | 0.887 | 0.893 |
| CKML Net/LGI [18] | 0.891 | 0.897 |
| Comprehensive CAD [14] | 0.91 | - |
| Expert A [14] | 0.94 | - |
| Expert B [14] | 0.92 | - |
| **Zoom-in-Net** | **0.957** | **0.911** |

**Table 3.** AUC for referable/nonreferable

| Method | AUC | Acc. |
|---|---|---|
| Splat feature/kNN [17] | 0.870 | - |
| VNXK/LGI [18] | 0.870 | 0.871 |
| CKML Net/LGI [18] | 0.862 | 0.858 |
| Comprehensive CAD [14] | 0.876 | - |
| Expert A [14] | 0.922 | - |
| Expert B [14] | 0.865 | - |
| **Zoom-in-Net** | **0.921** | **0.905** |

**Table 4.** AUC for normal/abnormal