Attention-based Skin Lesion Recognition

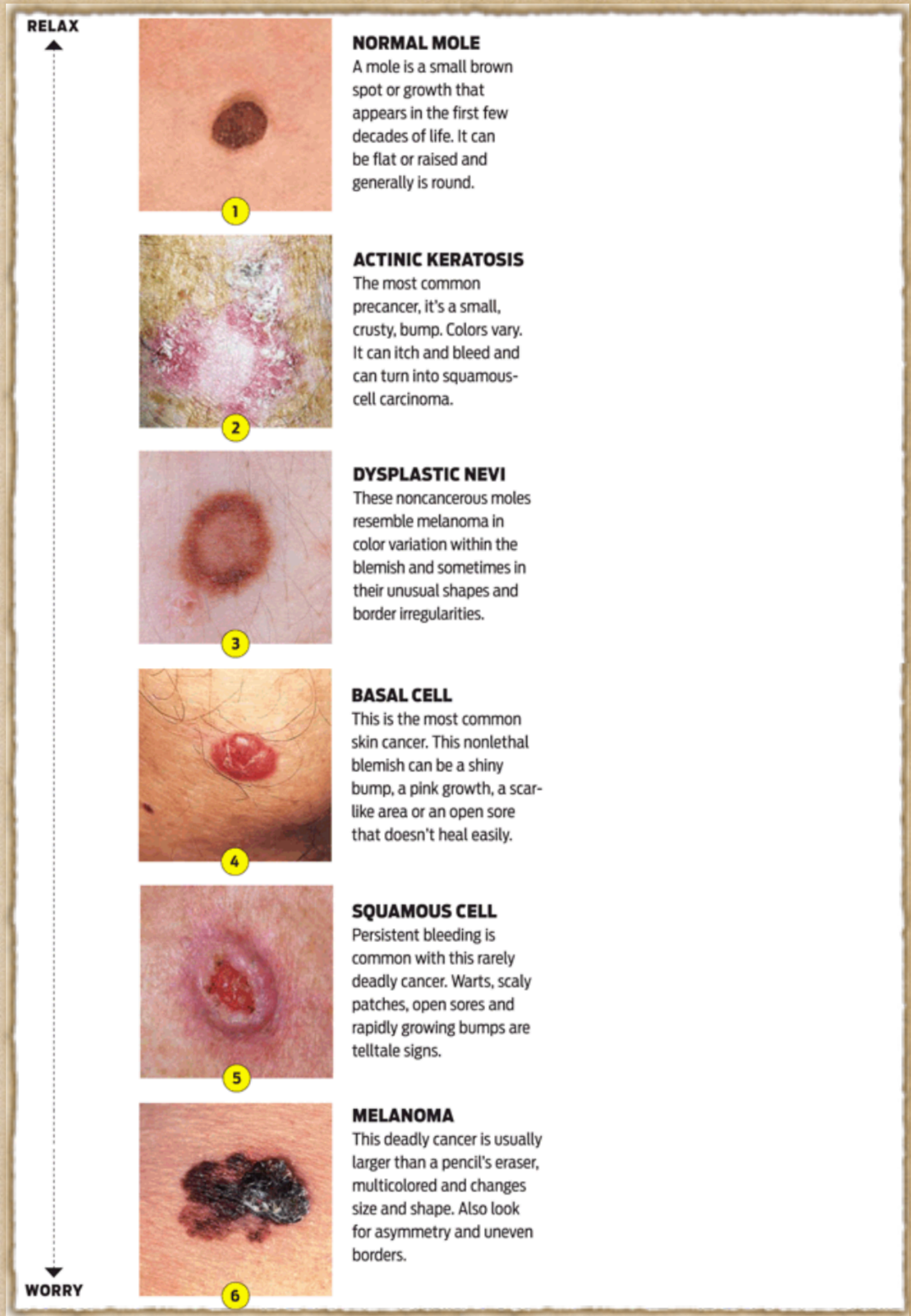
M.Sc Thesis Defense
School of Computing Science, Simon Fraser University
April 7, 2020

Yiqi Yan
B.Eng, Northwestern Polytechnical University

- **Background**
- Related Work
- Methodology
- Experiments
  - Setting up
  - Binary Classification
  - Multi-class Classification

Skin Cancer

- More than 2 people die of skin cancer in the U.S. every hour*
- When detected early, the 5-year survival rate for melanoma is 99 percent*



* Data source: Skin Cancer Foundation https://www.skincancer.org/skin-cancer-information/skin-cancer-facts/

Dermoscopy

- Non-invasive diagnosis;
- Improves diagnostic accuracy compared to standard photography;
- Portable devices are available;

- Background
- <u>Related Work</u>
- Methodology
- Experiments
  - Setting up
  - Binary Classification
  - Multi-class Classification

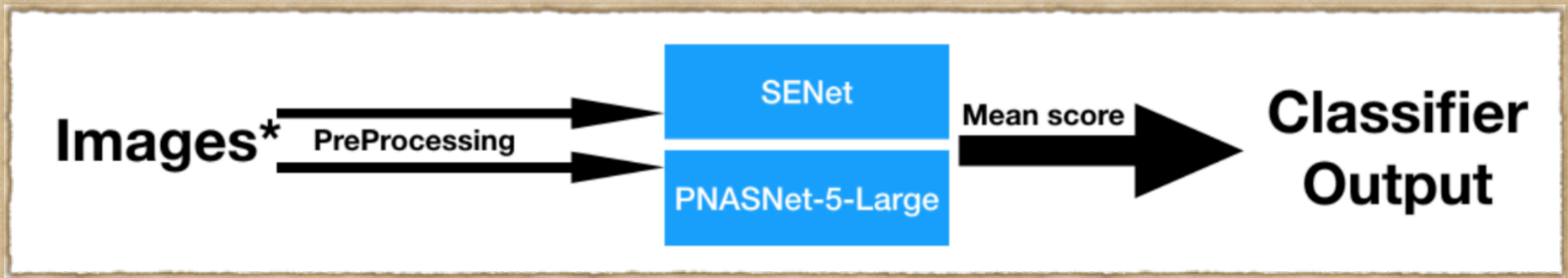
Related Work

What have been done

- Network/feature ensembles;
- Segmentation-guided classification;
- Interpreting results.

Network Ensembles

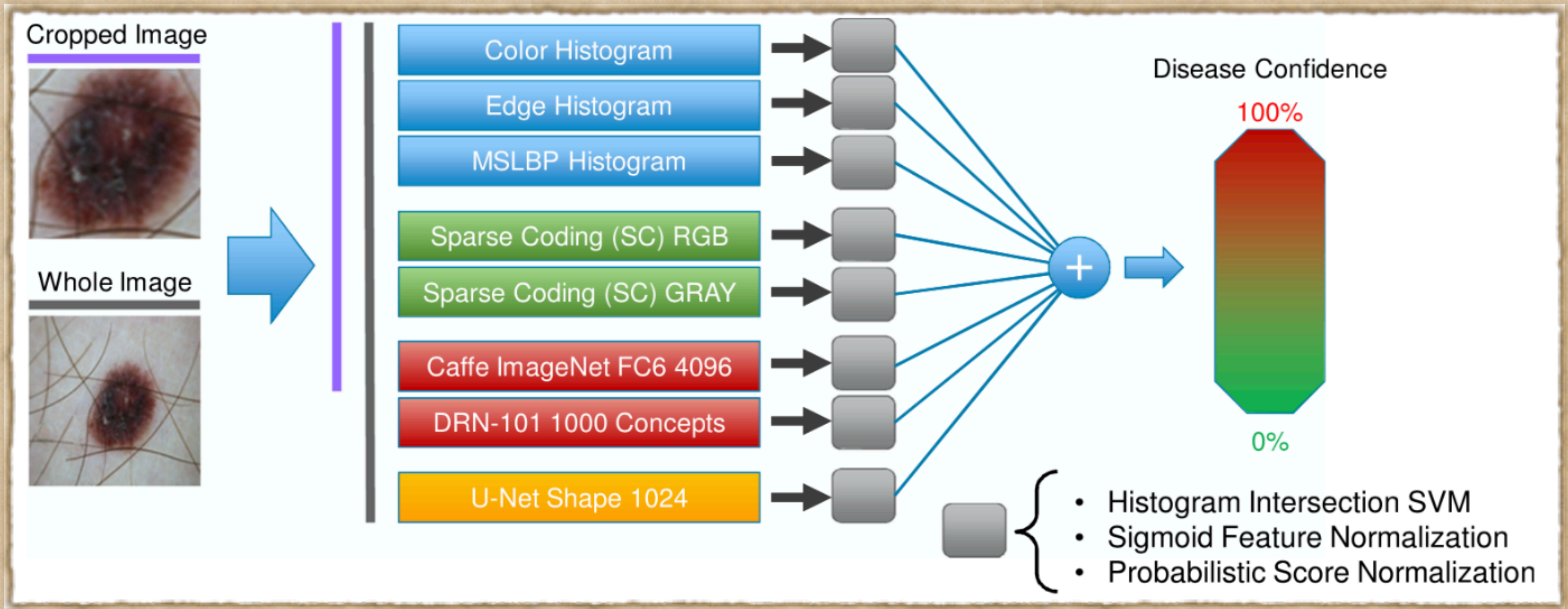
- High training cost;
- Coupling models: hard to tune



Zhuang et al. Skin lesion analysis towards melanoma detection using deep neural network ensemble. MICCAI, 2018.

Feature Ensembles

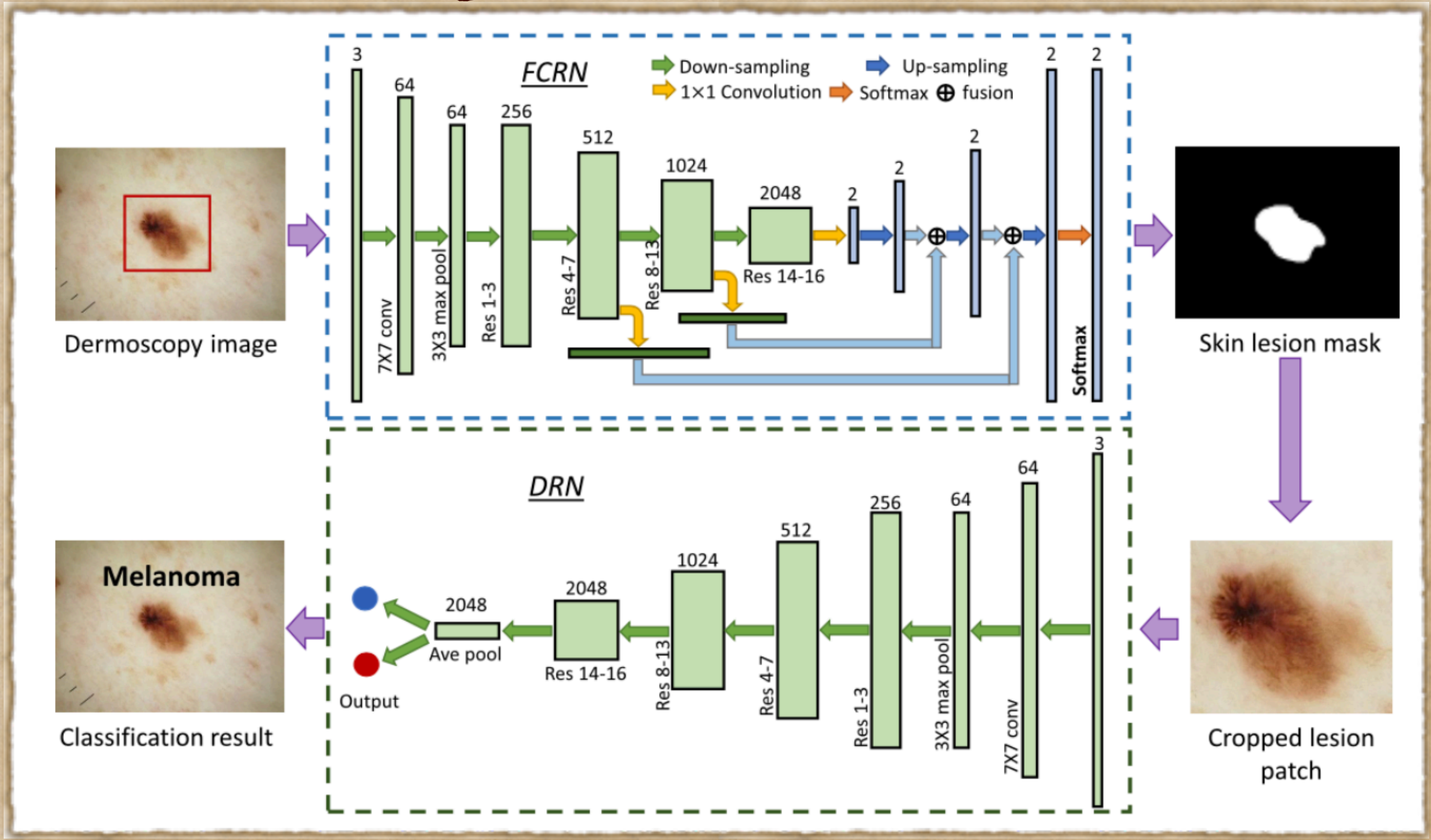
- Hand-crafted features: tricky to design;
- Deep features: requiring pre-training;
- Coupling features: hard to tune.



Codella et al. IBM Journal of Research and Development, 2017.

Segmentation-guided Classification - Sequential

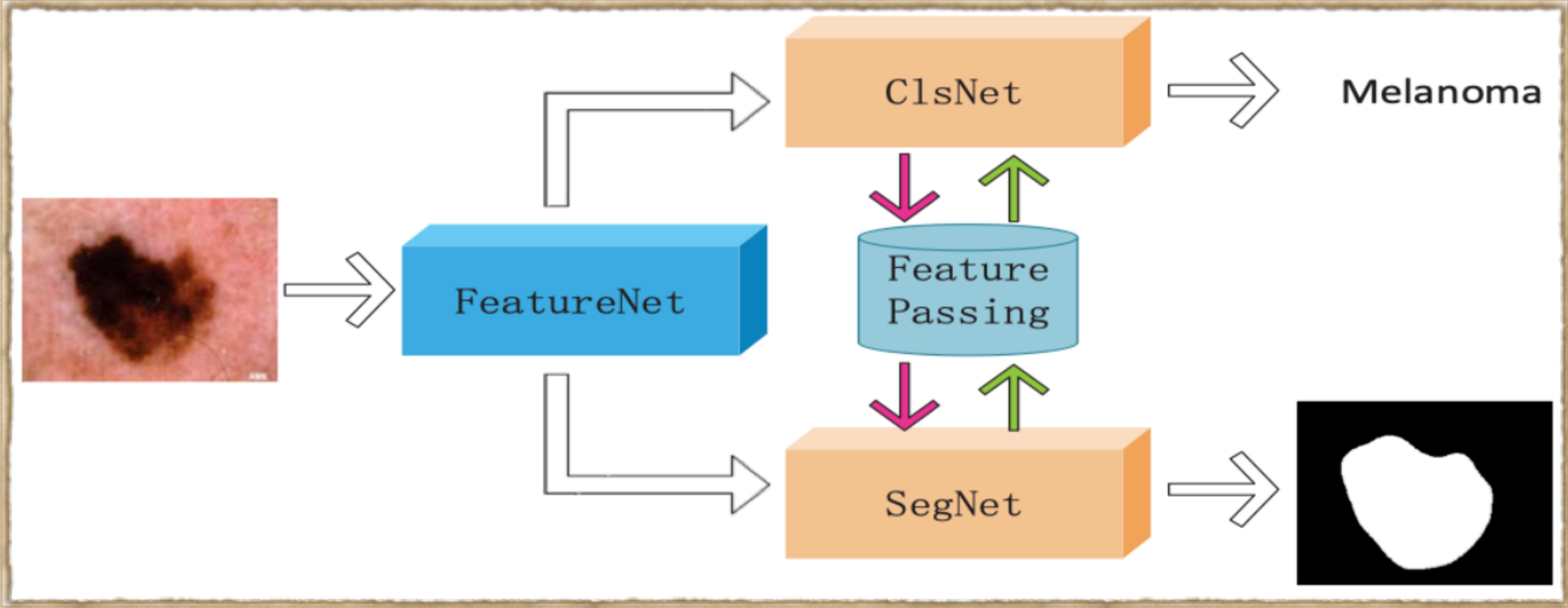
- Requiring accurate and complete pixel-level annotations;
- Relying much on the performance of the segmentation network;
- Not end-to-end training.



Yu et al. IEEE Transactions on Medical Imaging, 2017.

Segmentation-guided Classification - Parallel

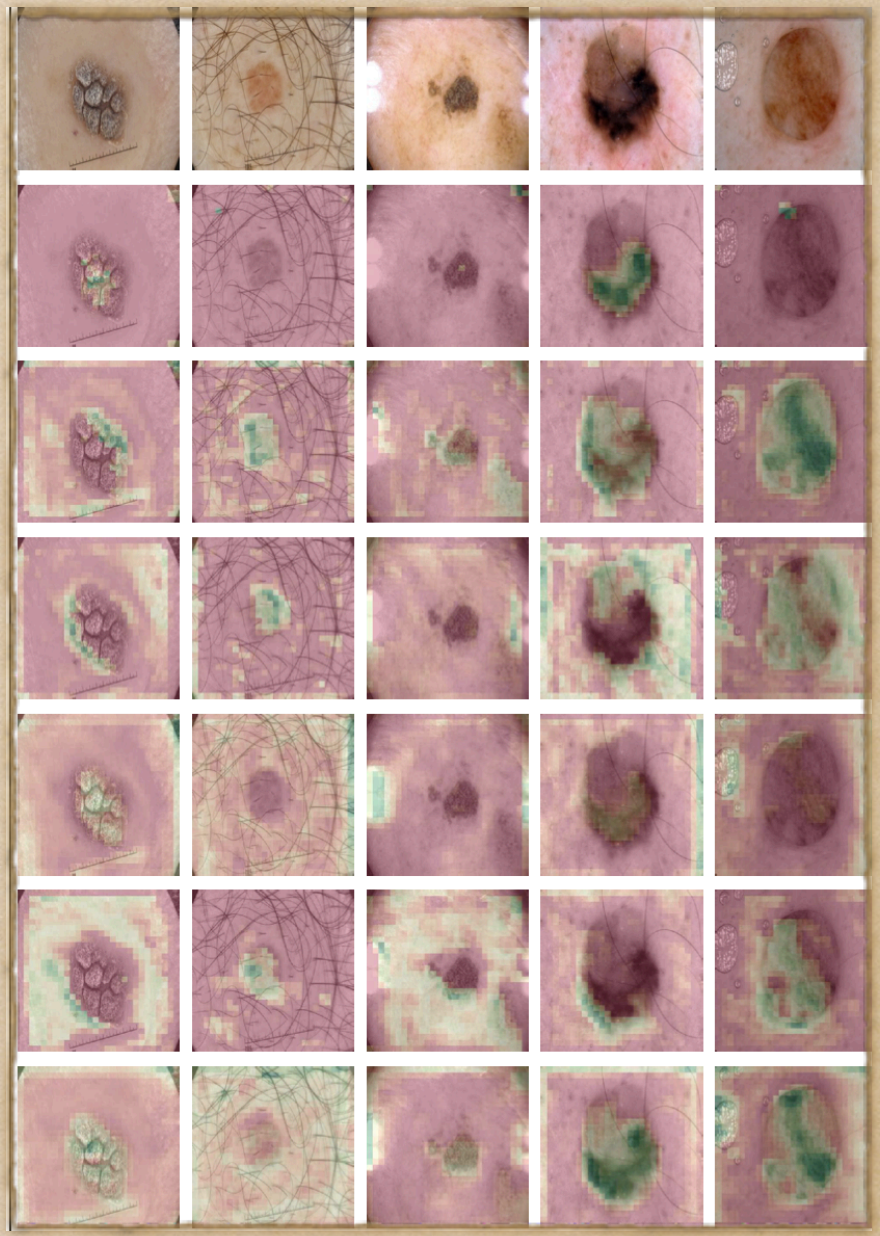
- Requiring accurate pixel-level annotations;
- The performance of the segmentation network affects classification accuracy.



Chen et al. A multi-task framework with feature passing module for skin lesion classification and segmentation. ISBI, 2018.

Visual Interpretability - Feature Map Visualization

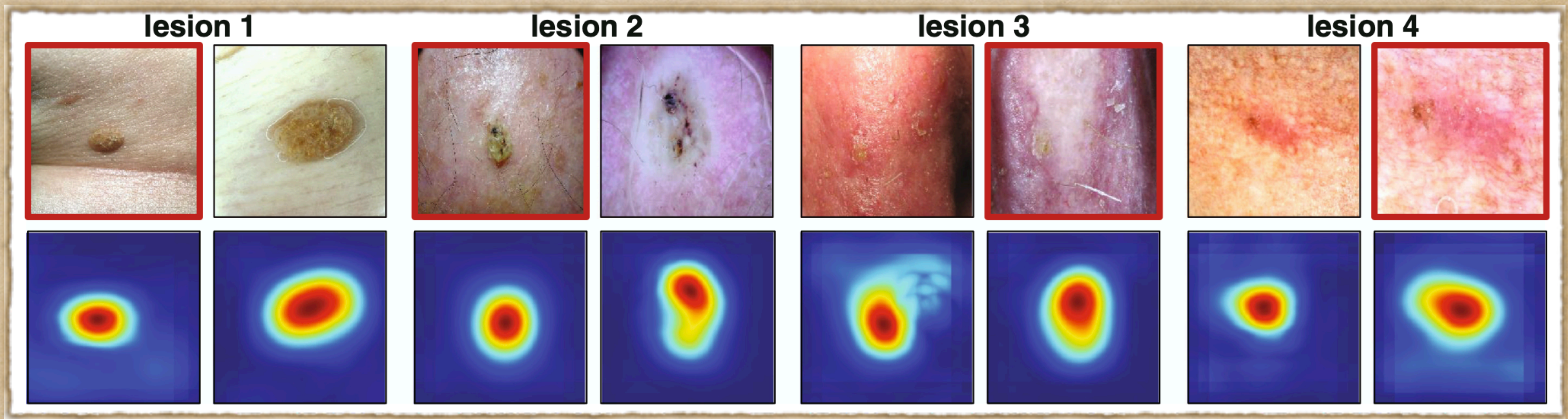
- Post hoc analysis based on fully trained models;
- Experimental hypothesis on what the feature seems to focus on;
- Interpretability only; not helping with classification performance.



Molle et al. MICCAI Workshop, 2018.

Visual Interpretability - Class Activation Map

- Post-processing based on fully trained models;

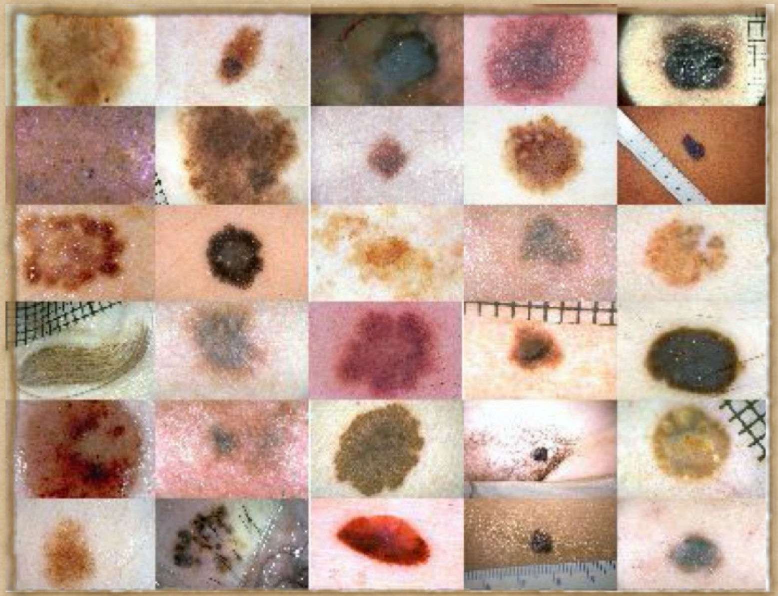


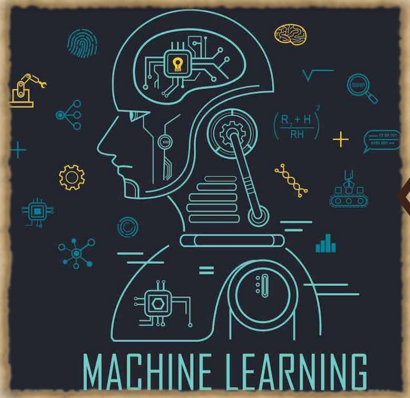
Ge et al. MICCAI, 2017

What can be improved

- End-to-end training; no complex ensembles or post-processing;
- Flexibility of applying pixel-level annotations
  - Using them as attention prior
  - Plug-in attention regularization term

- Background
- Related Work
- <u>Methodology</u>
- Experiments
  - Setting up
  - Binary Classification
  - Multi-class Classification

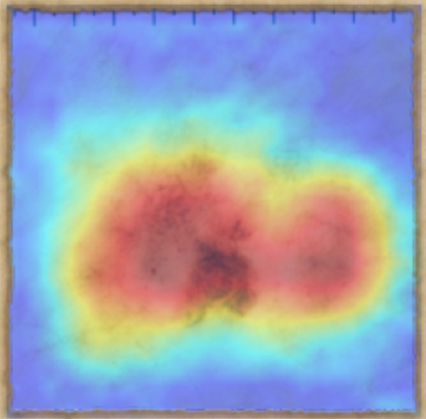


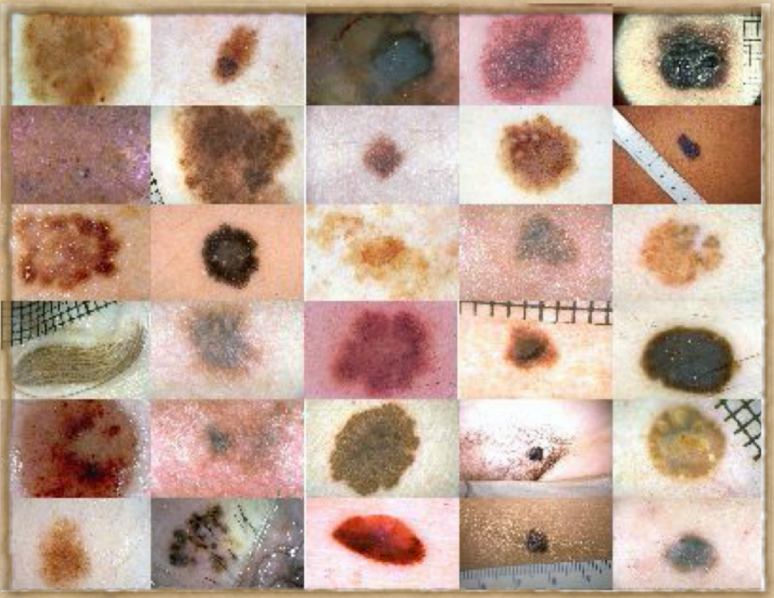


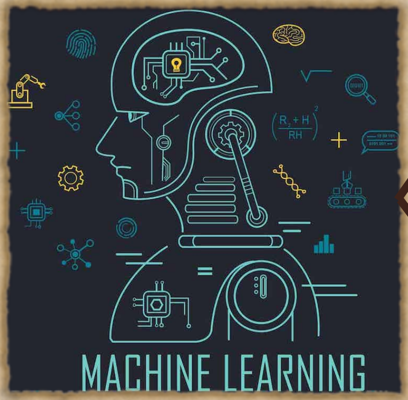
benign?

melanoma?

...





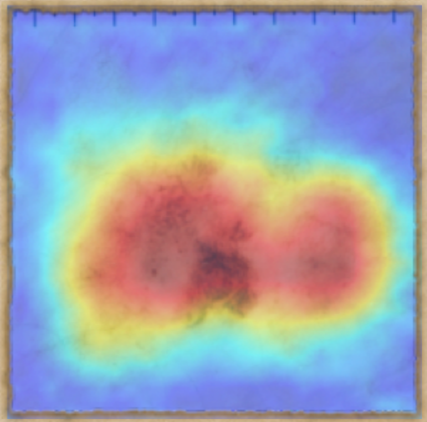


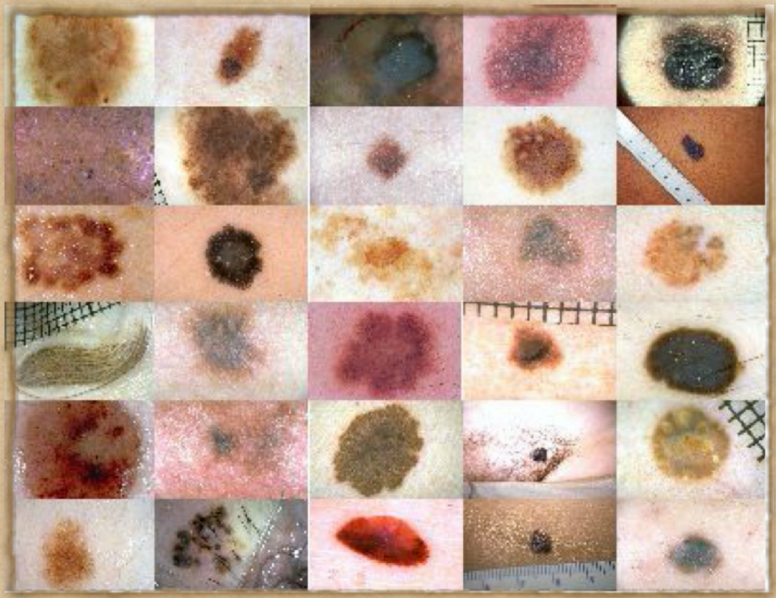
achieving better accuracy

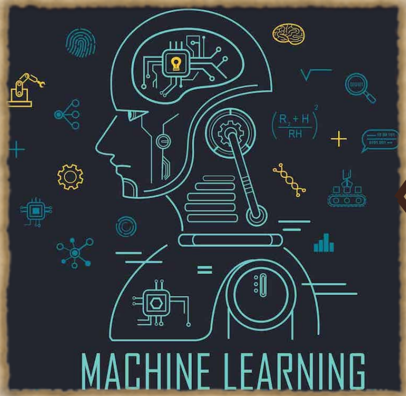
benign?

melanoma?

...







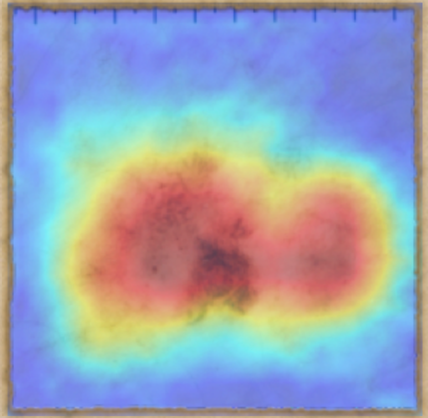
achieving better accuracy

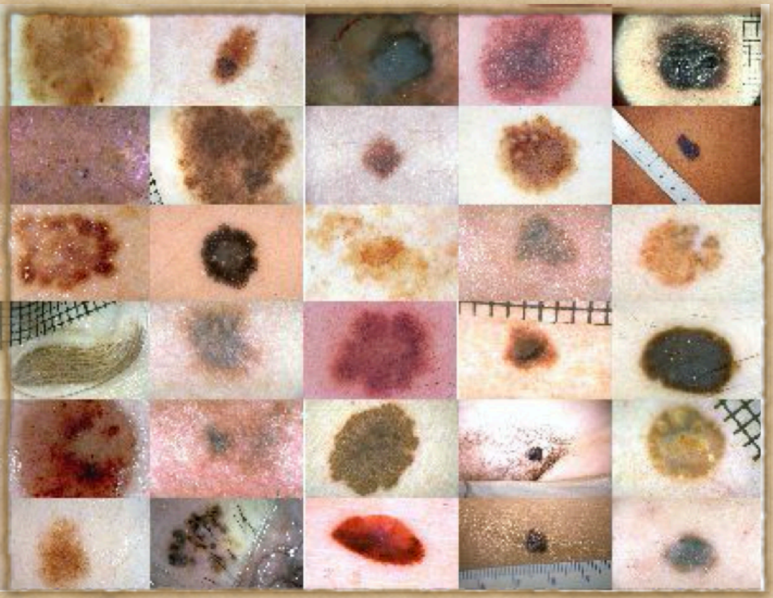
benign?

melanoma?

...

getting interpretable results



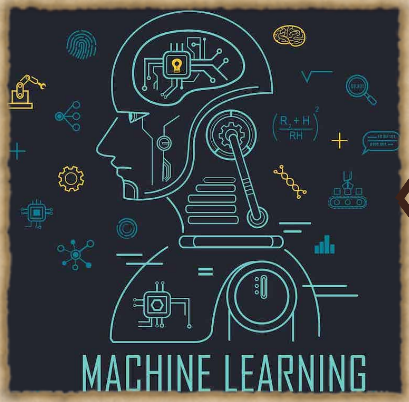


achieving better accuracy

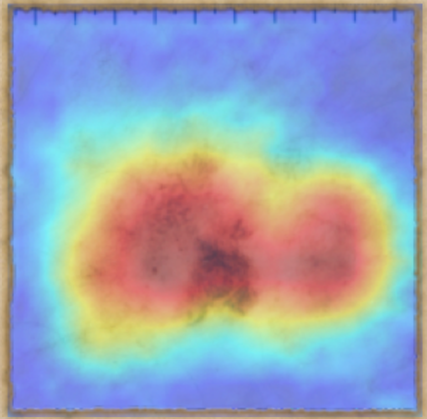
benign?

melanoma?

...

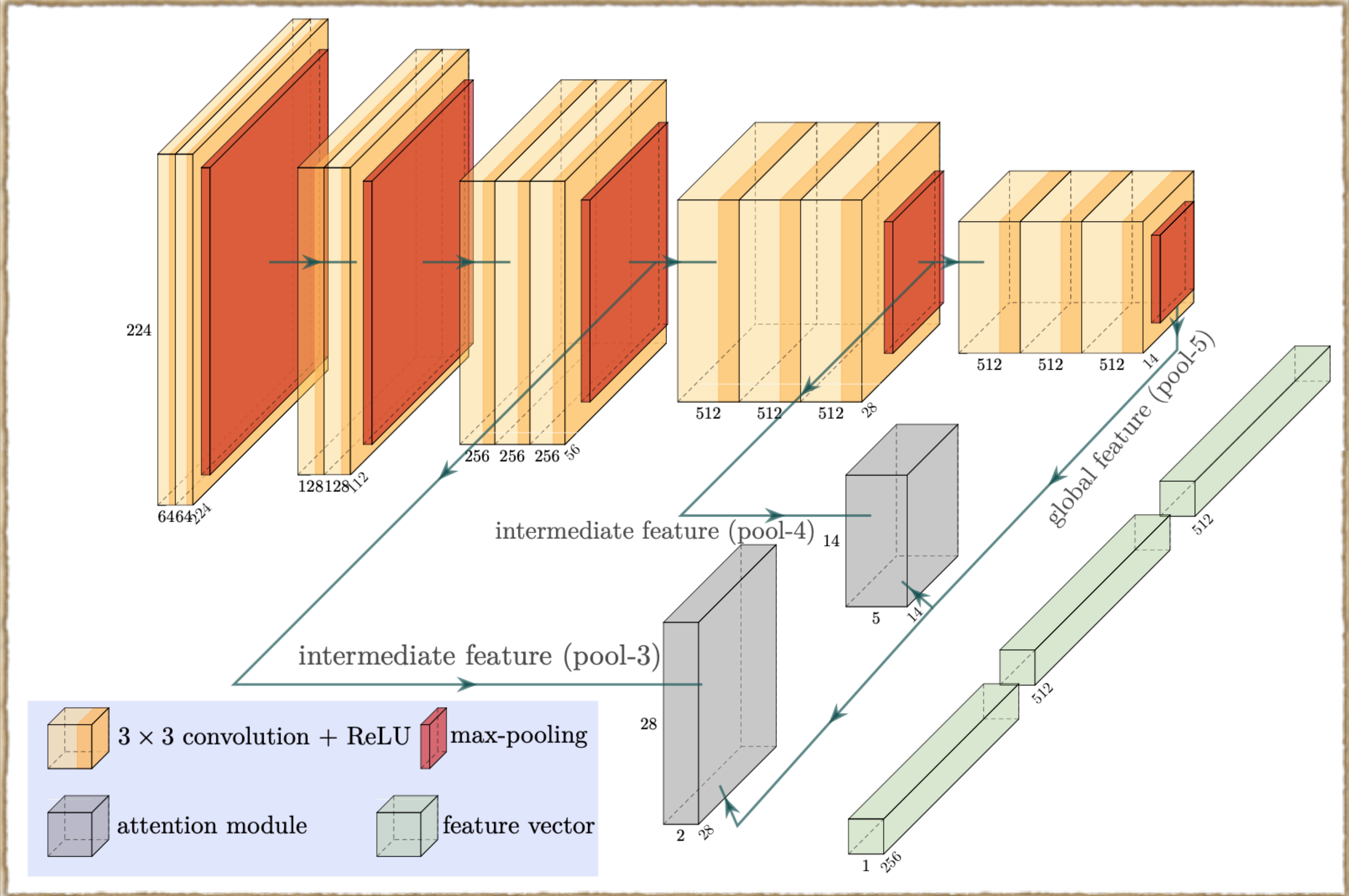


getting interpretable results

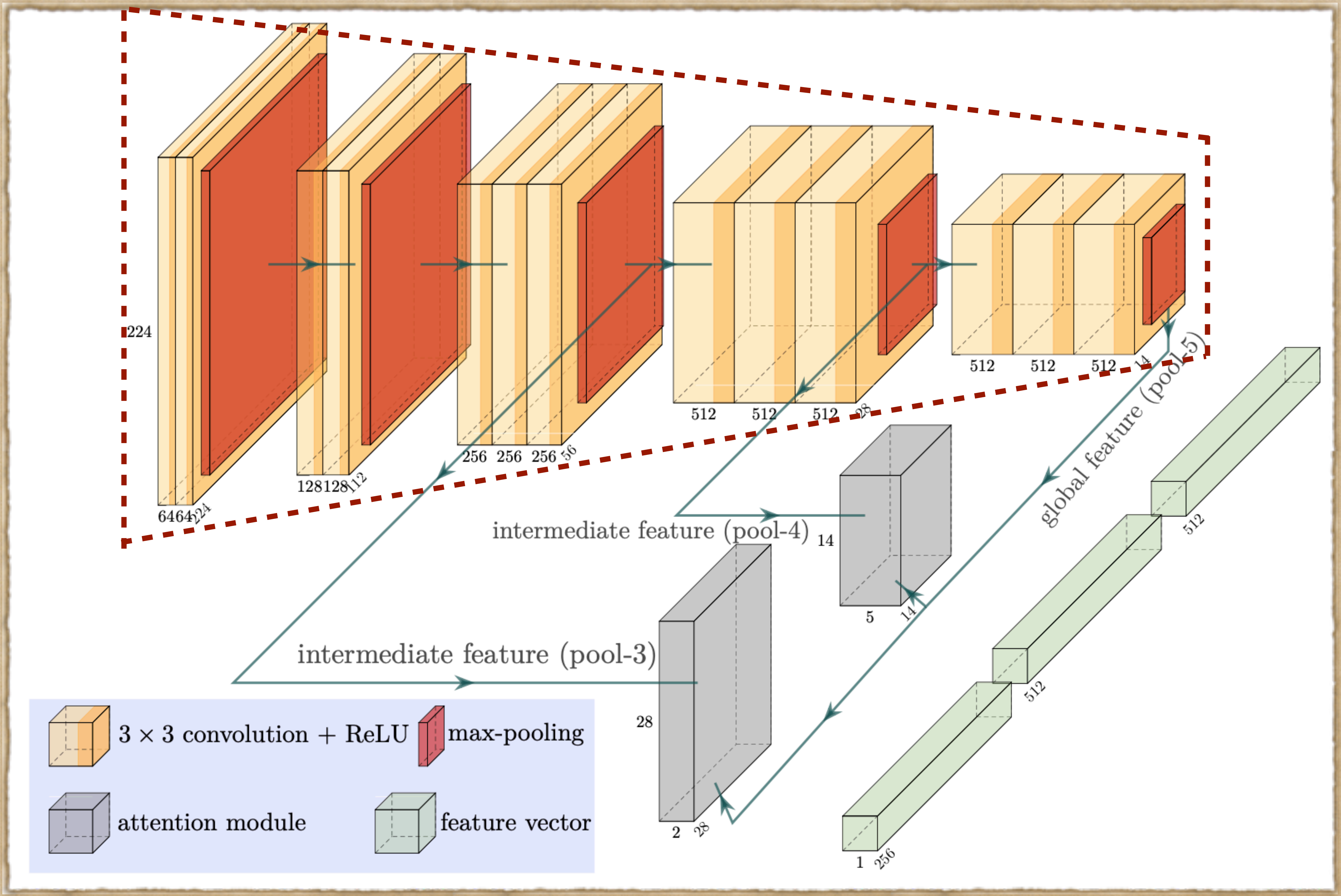


unified model: end-to-end training

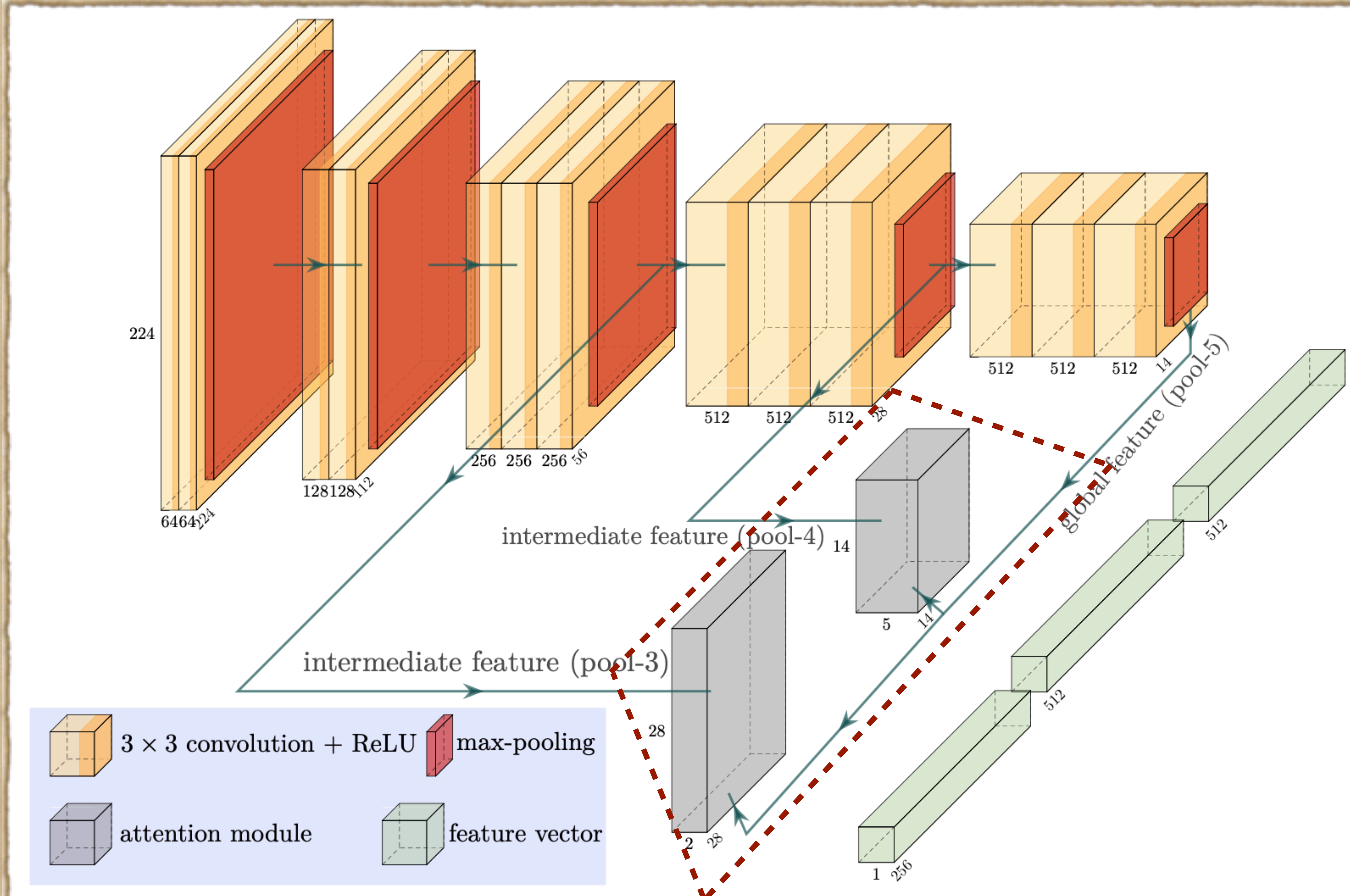
Overall Architecture



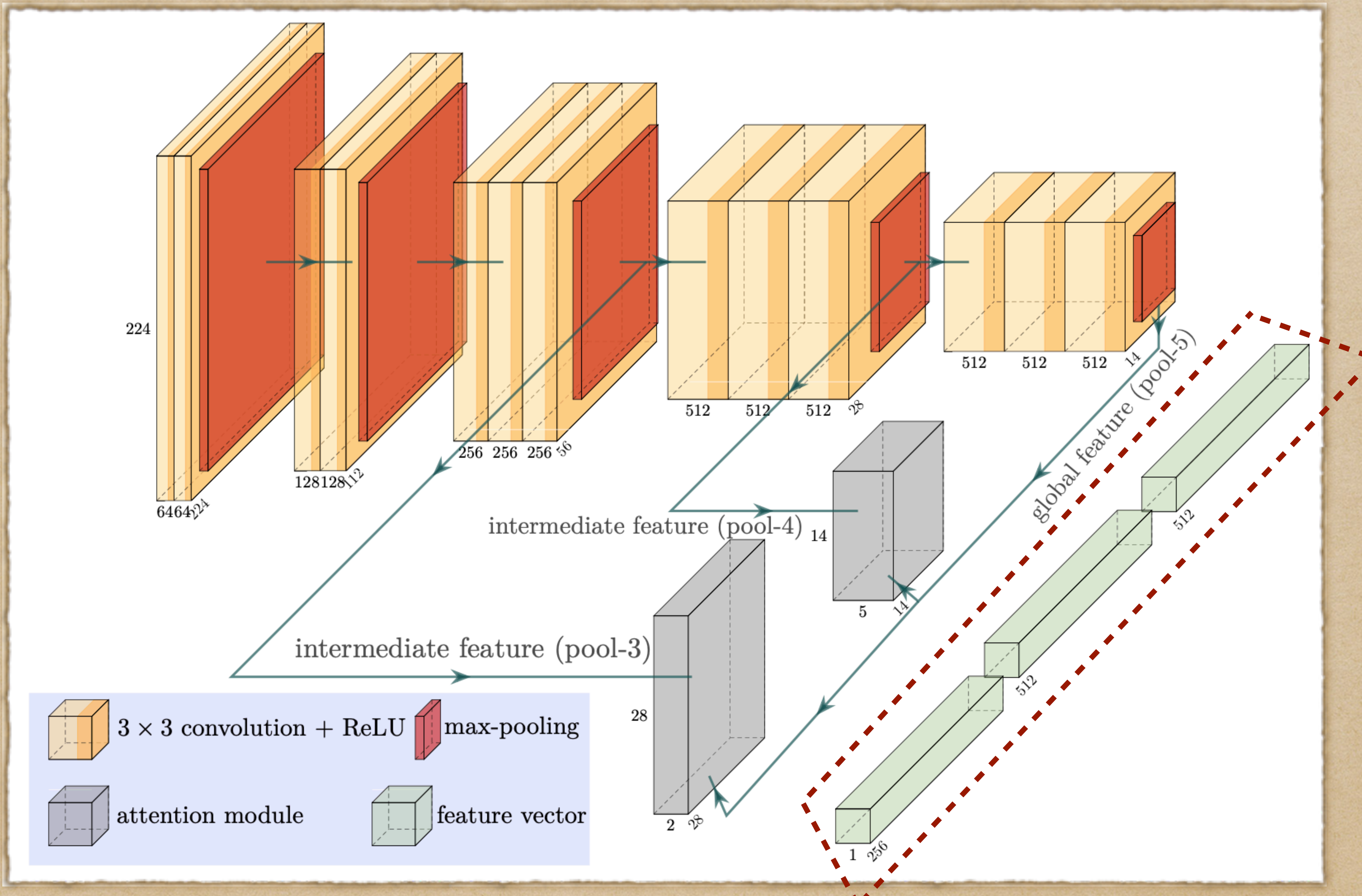
Backbone: VGG-16 (without dense layers)

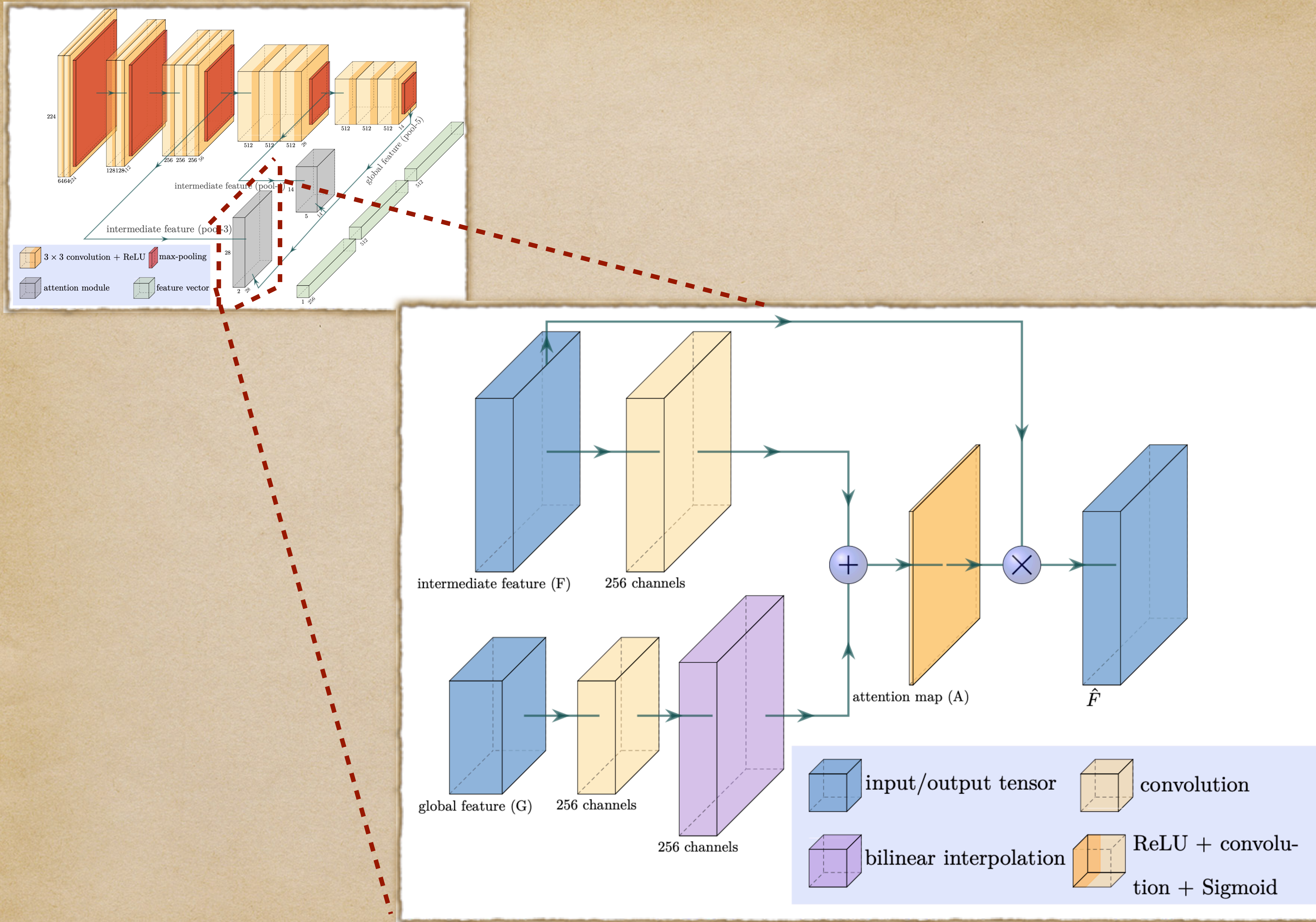


Attention Modules



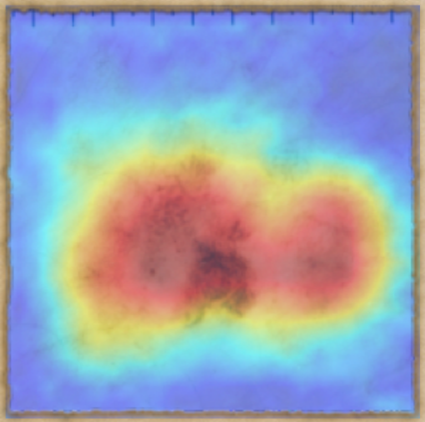
Global Average Pooling




224
512
intermediate feature (pool-3)
global feature (pool-5)
3 × 3 convolution + ReLU
max-pooling
attention module
feature vector
intermediate feature (F)
256 channels
global feature (G)
256 channels
256 channels
attention map (A)
$\hat{F}$
input/output tensor
convolution
bilinear interpolation
ReLU + convolution + Sigmoid

Attention Regularization

$$L_D(A, \overline{A}) = 1 - D(A, \overline{A}) = 1 - \frac{2 \cdot \sum_{i=1}^{n} \left(a_i \cdot \bar{a}_i\right)}{\sum_{i=1}^{n} \left(a_i + \bar{a}_i\right)}$$



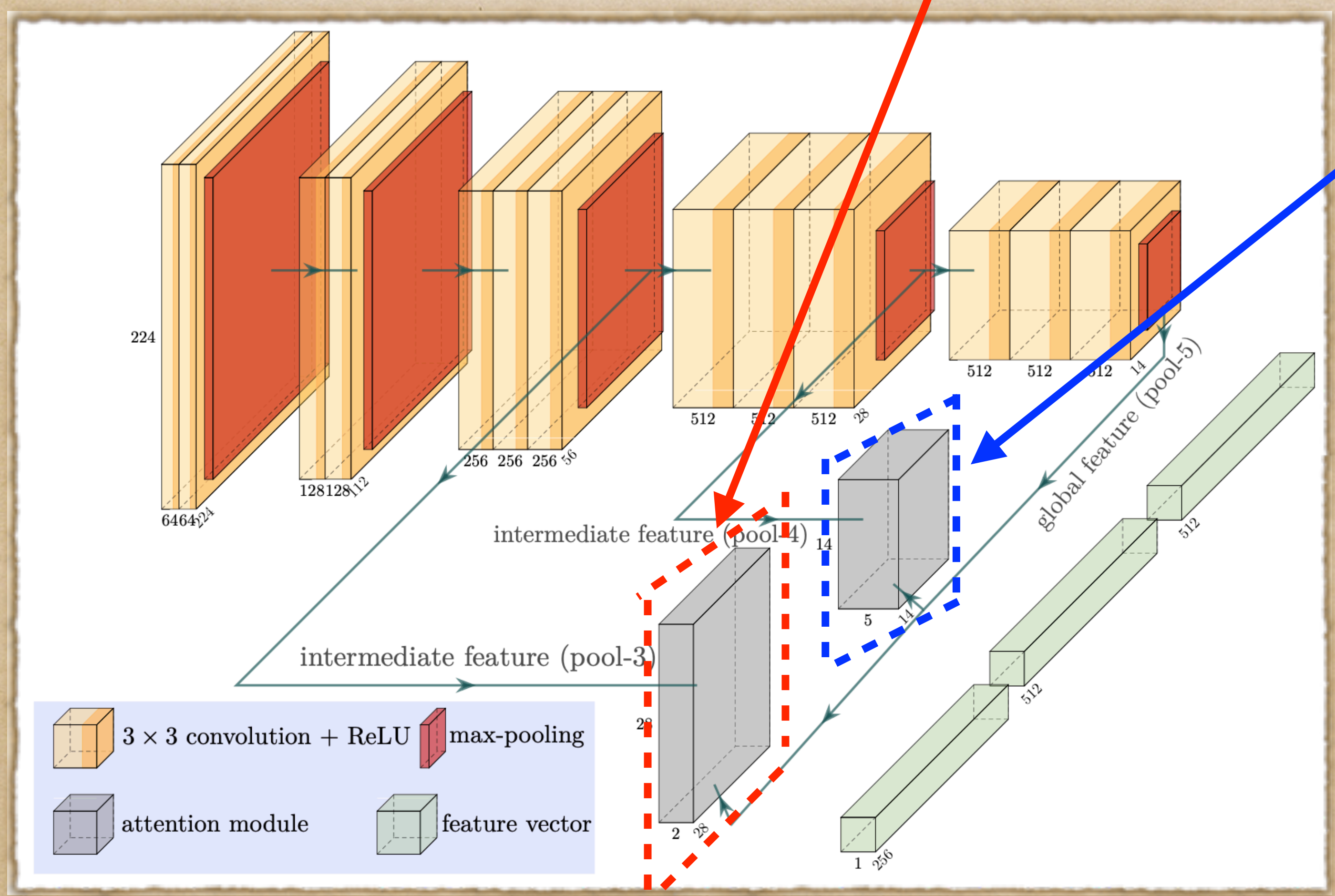
Lesion Segmentation

OR

Dermoscopic Features

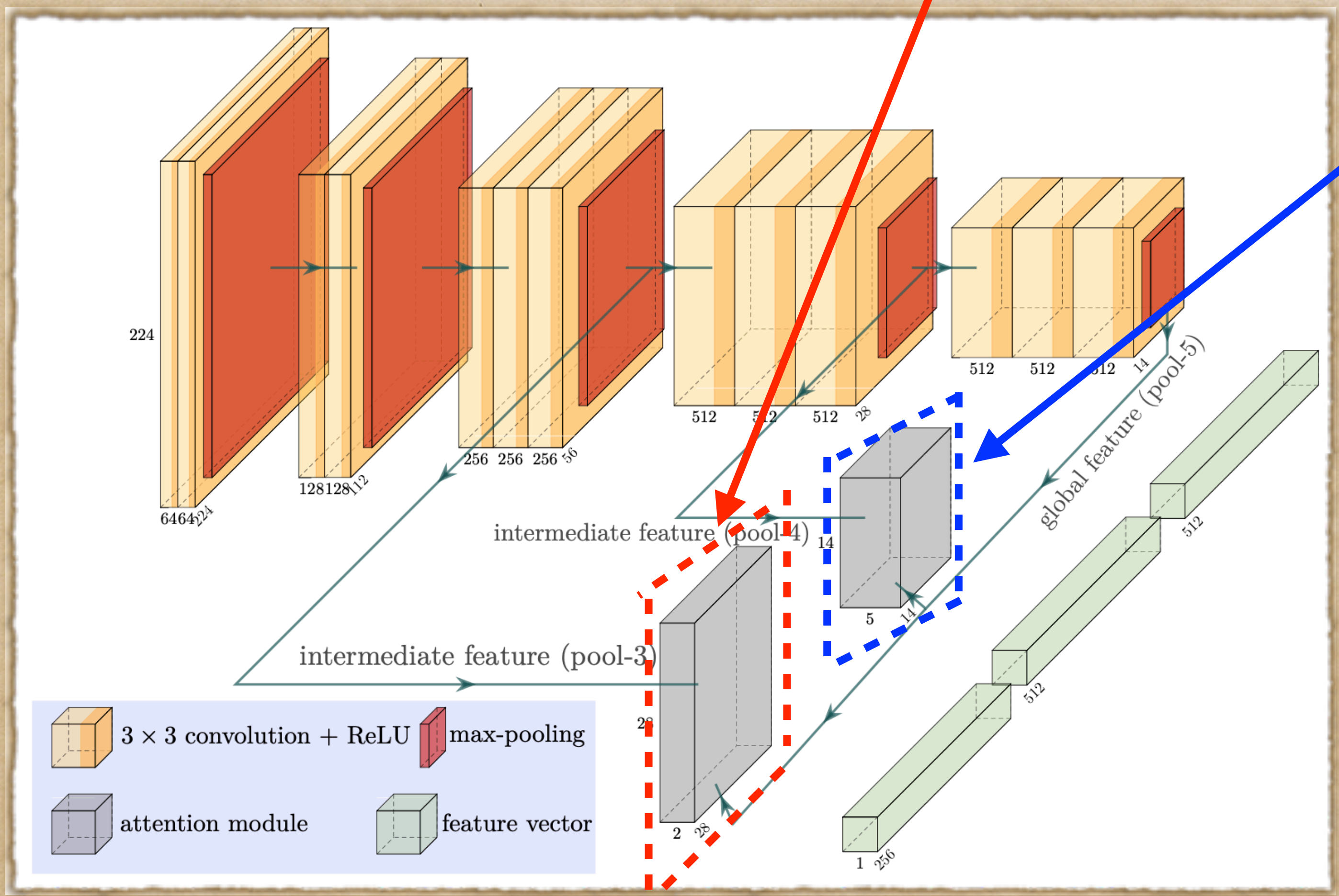
Complete Loss Function

$$L = L_{\text{focal}} + \lambda_1 L_D\left(A^{(3)}, \overline{A}^{(3)}\right) + \lambda_2 L_D\left(A^{(4)}, \overline{A}^{(4)}\right)$$



Complete Loss Function

$$L = L_{\text{focal}} + \underbrace{\lambda_1 L_D\left(A^{(3)}, \overline{A}^{(3)}\right)} + \underbrace{\lambda_2 L_D\left(A^{(4)}, \overline{A}^{(4)}\right)}$$



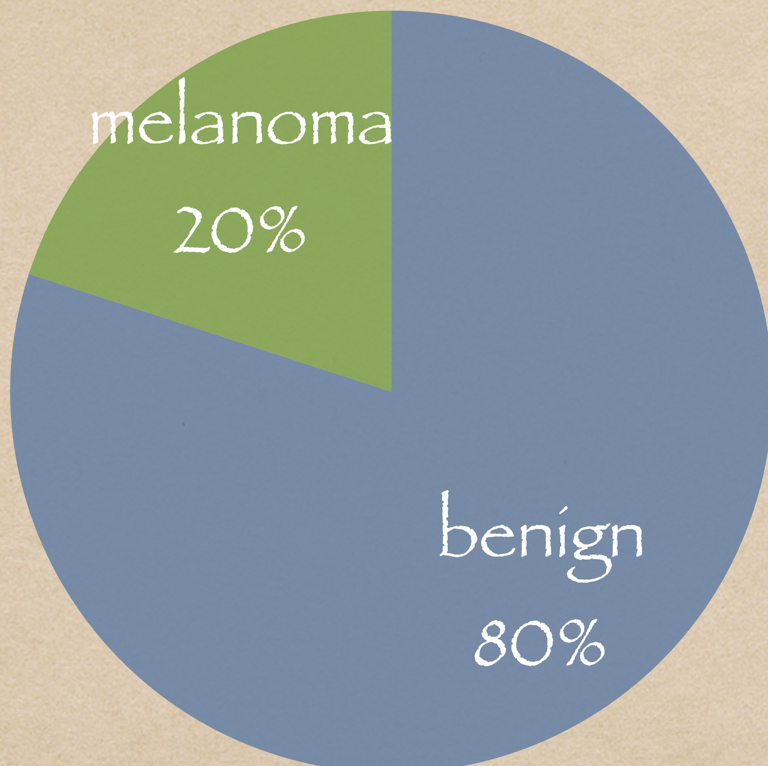
If pixel-level annotations are unavailable:

$$\lambda_1 = \lambda_2 = 0$$

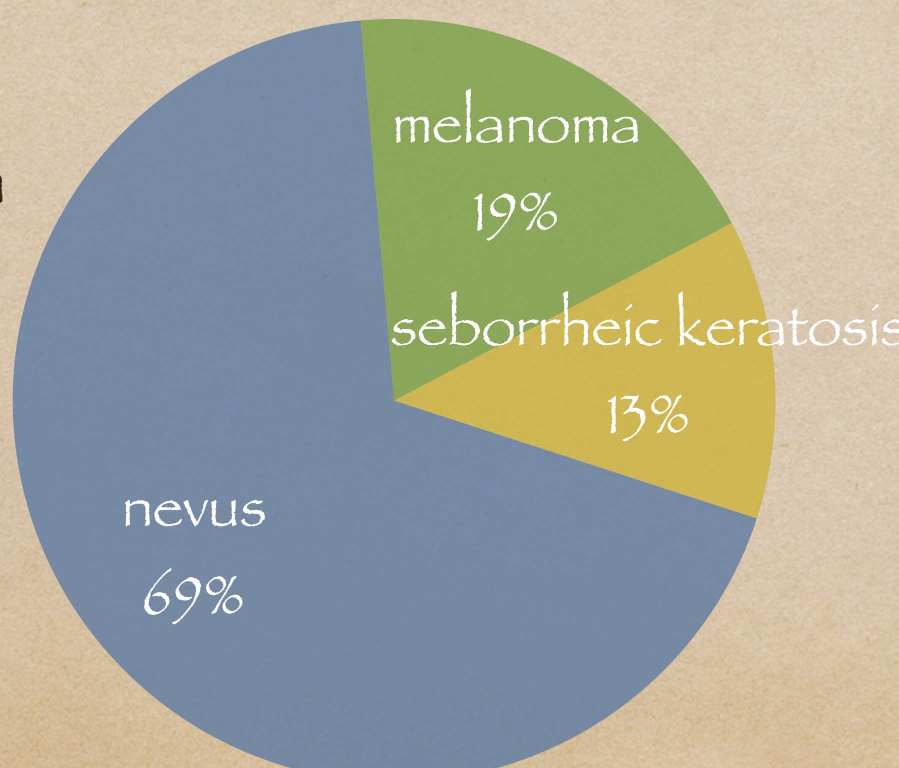
- Background
- Related Work
- Methodology
- Experiments
  - Setting up
  - Binary Classification
  - Multi-class Classification

Datasets

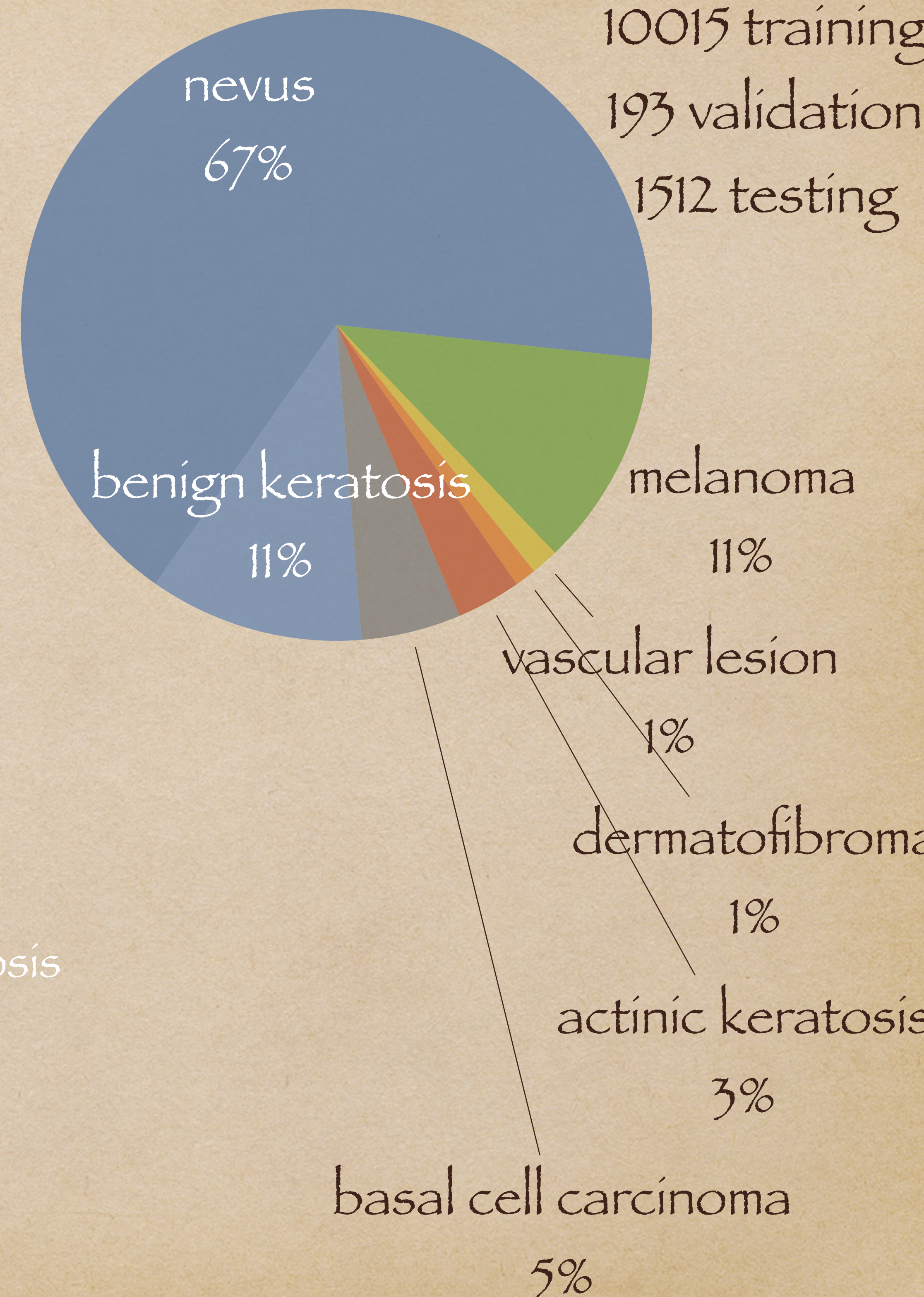
ISIC2016
900 training
379 testing



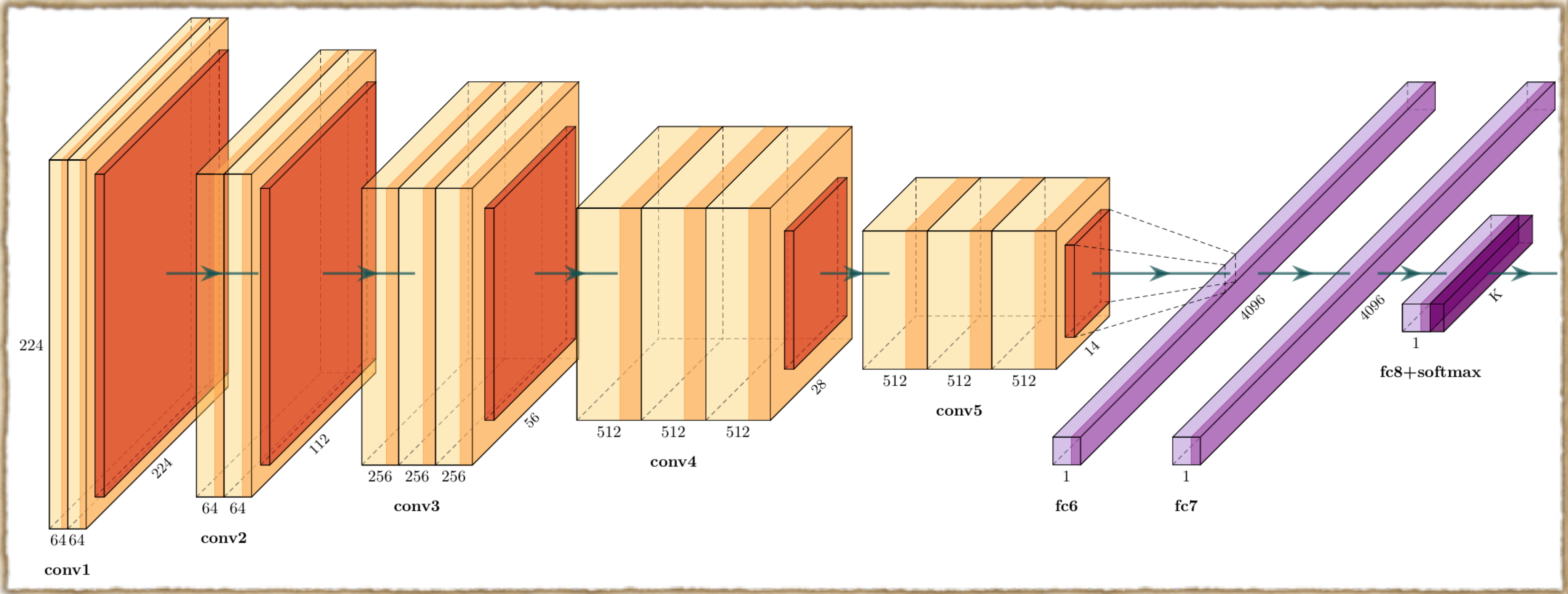
ISIC2017
200 training
150 validation
600 testing



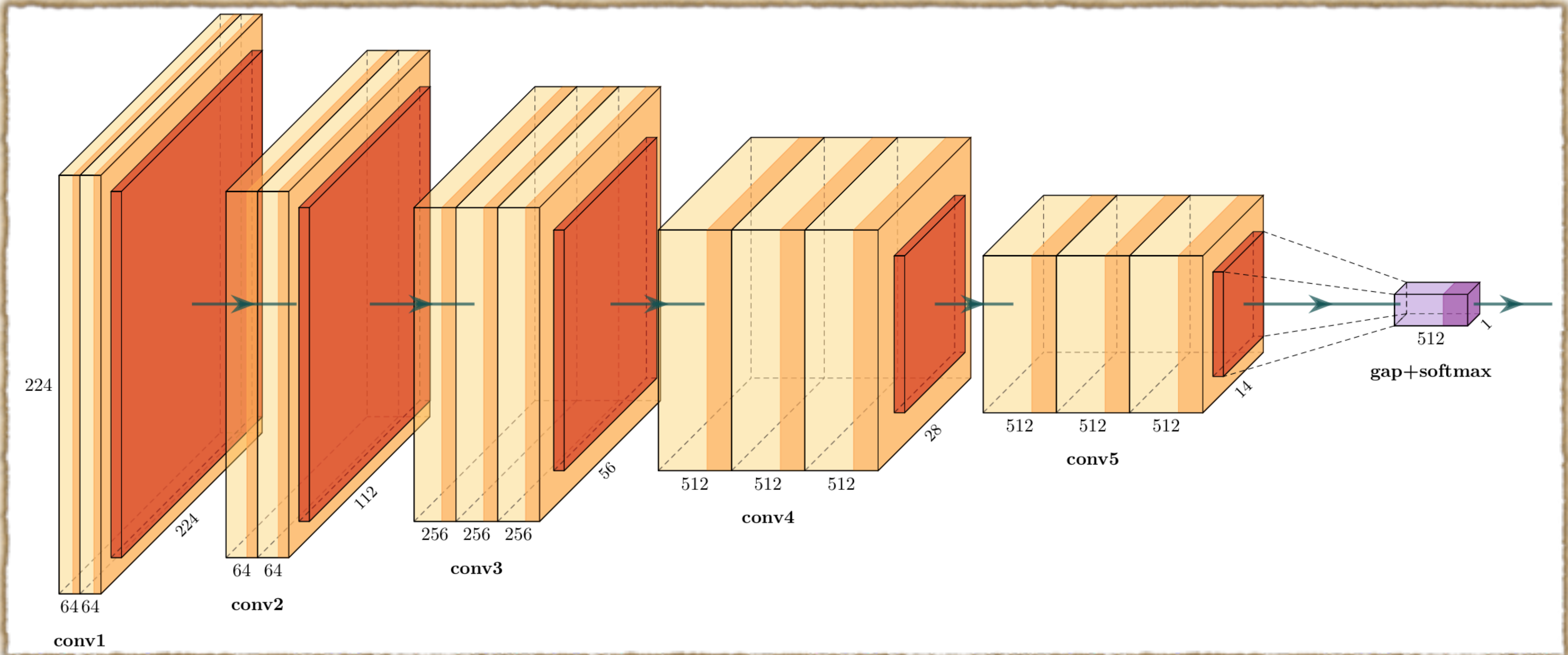
ISIC2018
10015 training
193 validation
1512 testing



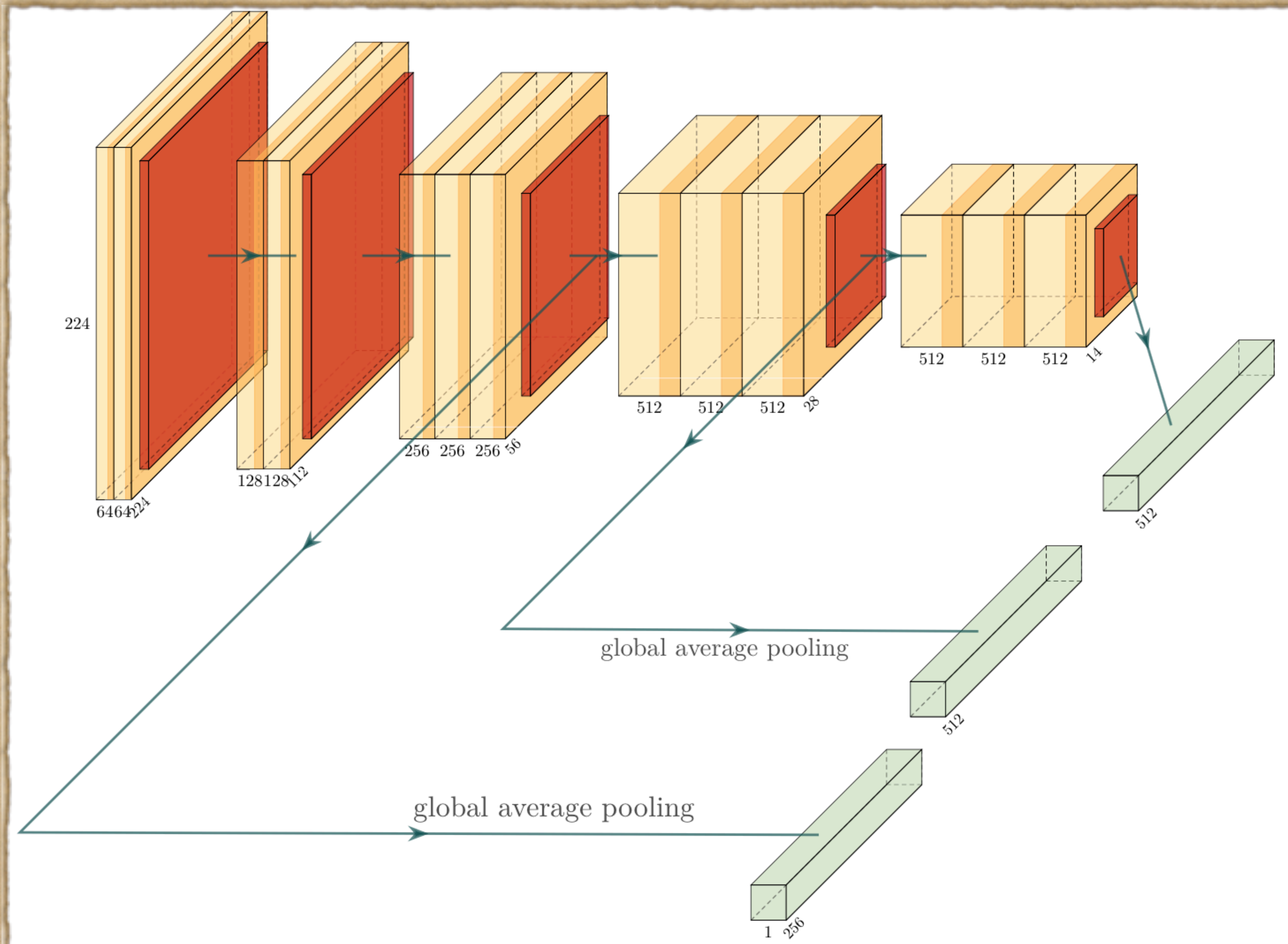
Baseline No.1 VGG-16



Baseline No.2 VGG-16-GAP



Baseline No.3 Mel-CNN

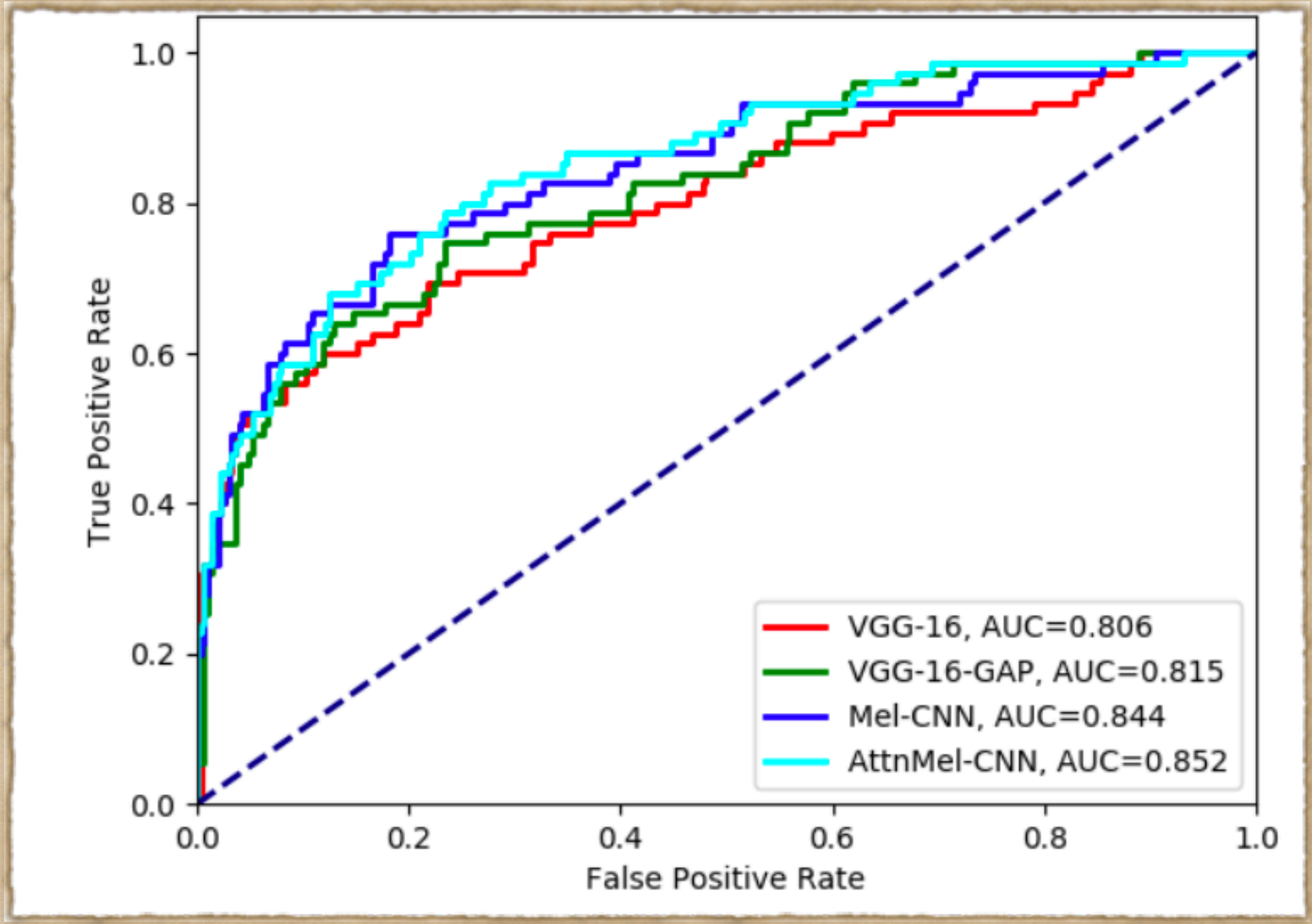


Network Training

- Software: PyTorch 1.0;
- Hardware: Nvidia GeForce GTX 1080 Ti
- Backbone network is initialized with ImageNet pre-trained parameters;
- Stochastic gradient descent with momentum; 50 epochs
- The initial learning rate is 0.01 and is decayed by 0.1 every 10 epochs;

- Background
- Related Work
- Methodology
- Experiments
    - Setting up
    - Binary Classification
    - Multi-class Classification

Results on ISIC2016



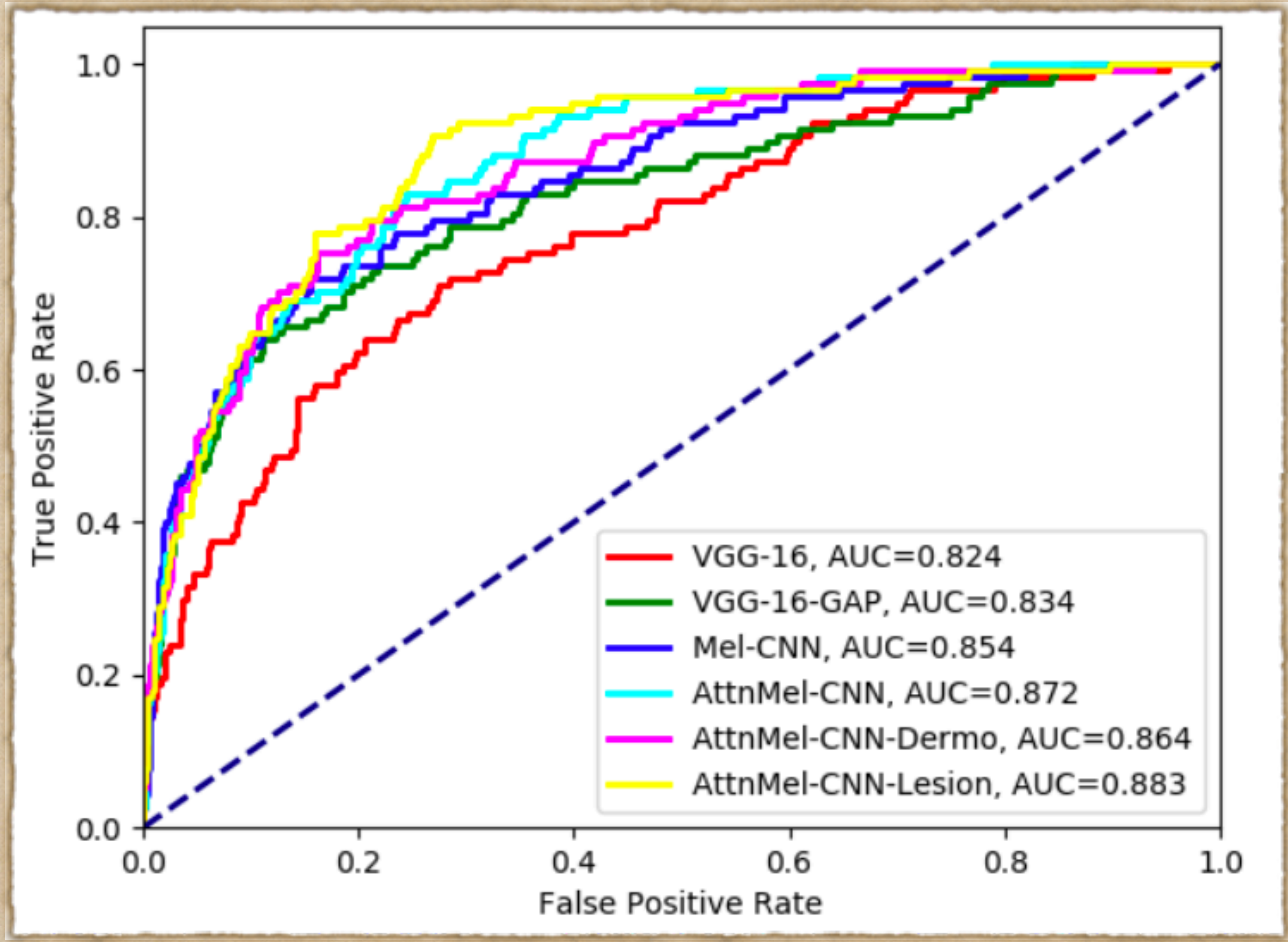
Results on ISIC2016

		AP	AUC	Lesion	Interp	Ensemble
#1	Lequan et al. [52]	0.637	0.804	✓	✗	✗
#2	Codella et al. [8]	0.596	0.808	✗	✗	✓
#3	Yu et al. [53, 54]	*0.685*	**0.852**	✗	✗	✓
#4	VGG-16	0.602	0.806	✗	✗	✗
#5	VGG-16-GAP	0.635	0.815	✗	✓	✗
#6	Mel-CNN	0.664	*0.844*	✗	✗	✗
#7	**AttnMel-CNN**	**0.693**	**0.852**	✗	✓	✗

Results on ISIC2016

		AP	AUC	Lesion	Interp	Ensemble
#1	Lequan et al. [52]	0.637	0.804	✓	✗	✗
#2	Codella et al. [8]	0.596	0.808	✗	✗	✓
#3	Yu et al. [53, 54]	*0.685*	**0.852**	✗	✗	✓
#4	VGG-16	0.602	0.806	✗	✗	✗
#5	VGG-16-GAP	0.635	0.815	✗	✓	✗
#6	Mel-CNN	0.664	*0.844*	✗	✗	✗
#7	**AttnMel-CNN**	**0.693**	**0.852**	✗	✓	✗

Results on ISIC2017



Results on ISIC2017

		AP	AUC	Lesion	Dermo	Interp	Ensemble	External
#1	Winner 1 [29]	–	0.868	✗	✗	✗	✓	✓
#2	Winner 2 [9]	–	0.856	✓	✓	✗	✗	✓
#3	Winner 3 [31]	–	*0.874*	✗	✗	✗	✓	✓
#4	Harangi et al. [14]	–	0.836	✗	✗	✗	✓	✗
#5	Mahbod et al. [28]	–	*0.873*	✗	✗	✗	✓	✓
#6	VGG-16	0.600	0.824	✗	✗	✗	✗	✗
#7	VGG-16-GAP	0.627	0.834	✗	✗	✓	✗	✗
#8	Mel-CNN	0.653	0.854	✗	✗	✗	✗	✗
#9	**AttnMel-CNN**	0.655	0.872	✗	✗	✓	✗	✗
#10	**AttnMel-CNN-Dermo**	*0.665*	0.864	✗	✓	✓	✗	✗
#11	**AttnMel-CNN-Lesion**	**0.672**	**0.883**	✓	✗	✓	✗	✗
#12	AttnMel-CNN-Bkg	0.647	0.849	✓	✗	✓	✗	✗

Results on ISIC2017

		AP	AUC	Lesion	Dermo	Interp	Ensemble	External
#1	Winner 1 [29]	–	0.868	✗	✗	✗	✓	✓
#2	Winner 2 [9]	–	0.856	✓	✓	✗	✗	✓
#3	Winner 3 [31]	–	*0.874*	✗	✗	✗	✓	✓
#4	Harangi et al. [14]	–	0.836	✗	✗	✗	✓	✗
#5	Mahbod et al. [28]	–	*0.873*	✗	✗	✗	✓	✓
#6	VGG-16	0.600	0.824	✗	✗	✗	✗	✗
#7	VGG-16-GAP	0.627	0.834	✗	✗	✓	✗	✗
#8	Mel-CNN	0.653	0.854	✗	✗	✗	✗	✗
#9	**AttnMel-CNN**	0.655	0.872	✗	✗	✓	✗	✗
#10	**AttnMel-CNN-Dermo**	*0.665*	0.864	✗	✓	✓	✗	✗
#11	**AttnMel-CNN-Lesion**	**0.672**	**0.883**	✓	✗	✓	✗	✗
#12	AttnMel-CNN-Bkg	0.647	0.849	✓	✗	✓	✗	✗

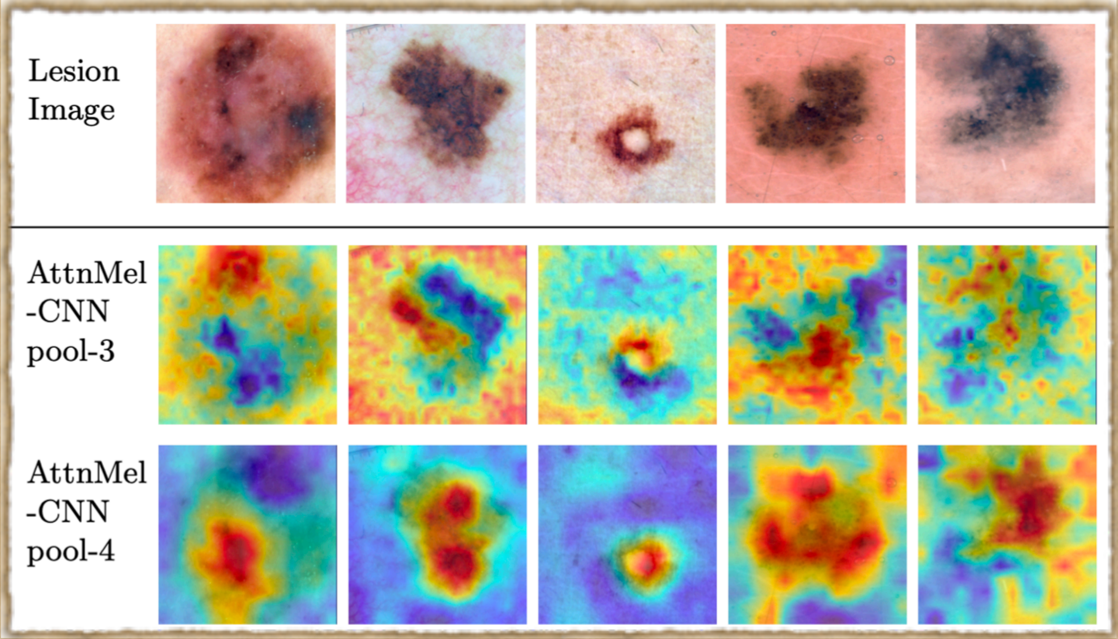
Results on ISIC2017

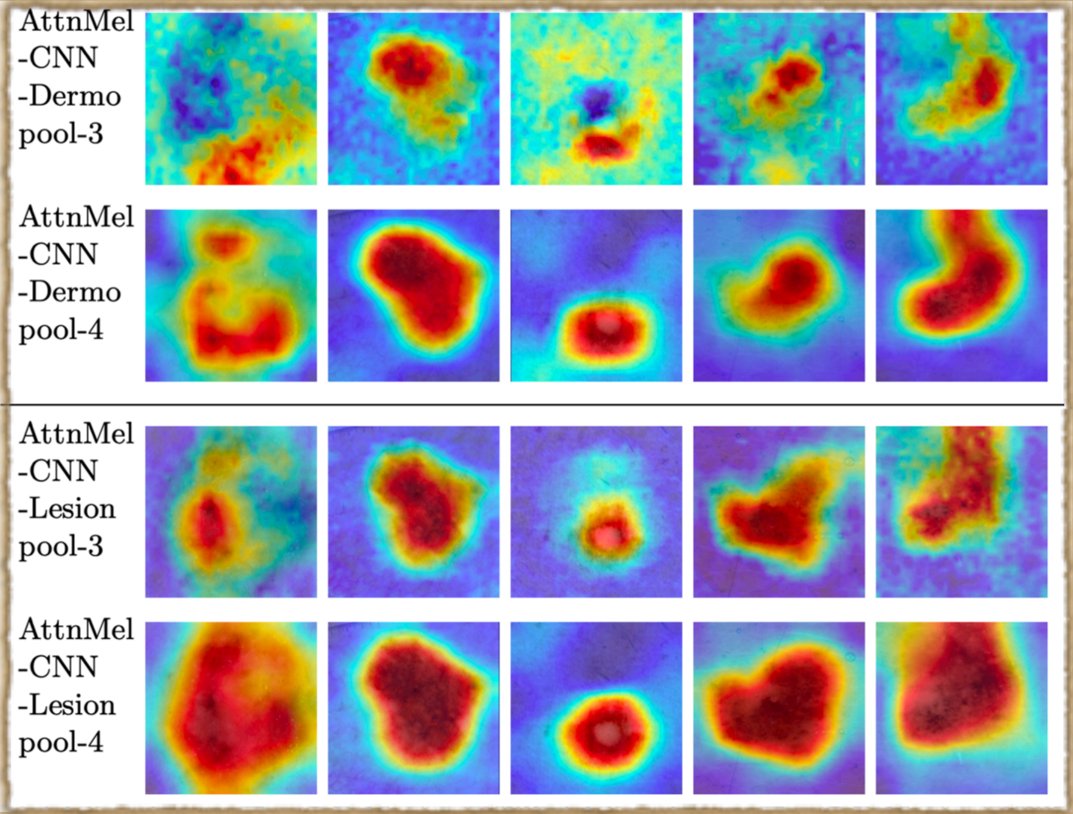
		AP	AUC	Lesion	Dermo	Interp	Ensemble	External
#1	Winner 1 [29]	–	0.868	✗	✗	✗	✓	✓
#2	Winner 2 [9]	–	0.856	✓	✓	✗	✗	✓
#3	Winner 3 [31]	–	*0.874*	✗	✗	✗	✓	✓
#4	Harangi et al. [14]	–	0.836	✗	✗	✗	✓	✗
#5	Mahbod et al. [28]	–	*0.873*	✗	✗	✗	✓	✓
#6	VGG-16	0.600	0.824	✗	✗	✗	✗	✗
#7	VGG-16-GAP	0.627	0.834	✗	✗	✓	✗	✗
#8	Mel-CNN	0.653	0.854	✗	✗	✗	✗	✗
#9	**AttnMel-CNN**	0.655	0.872	✗	✗	✓	✗	✗
#10	**AttnMel-CNN-Dermo**	*0.665*	0.864	✗	✓	✓	✗	✗
#11	**AttnMel-CNN-Lesion**	**0.672**	**0.883**	✓	✗	✓	✗	✗
#12	AttnMel-CNN-Bkg	0.647	0.849	✓	✗	✓	✗	✗

Results on ISIC2017

		AP	AUC	Lesion	Dermo	Interp	Ensemble	External
#1	Winner 1 [29]	–	0.868	✗	✗	✗	✓	✓
#2	Winner 2 [9]	–	0.856	✓	✓	✗	✗	✓
#3	Winner 3 [31]	–	*0.874*	✗	✗	✗	✓	✓
#4	Harangi et al. [14]	–	0.836	✗	✗	✗	✓	✗
#5	Mahbod et al. [28]	–	*0.873*	✗	✗	✗	✓	✓
#6	VGG-16	0.600	0.824	✗	✗	✗	✗	✗
#7	VGG-16-GAP	0.627	0.834	✗	✗	✓	✗	✗
#8	Mel-CNN	0.653	0.854	✗	✗	✗	✗	✗
#9	**AttnMel-CNN**	0.655	0.872	✗	✗	✓	✗	✗
#10	**AttnMel-CNN-Dermo**	*0.665*	0.864	✗	✓	✓	✗	✗
#11	**AttnMel-CNN-Lesion**	**0.672**	**0.883**	✓	✗	✓	✗	✗
#12	AttnMel-CNN-Bkg	0.647	0.849	✓	✗	✓	✗	✗

Results on ISIC2017

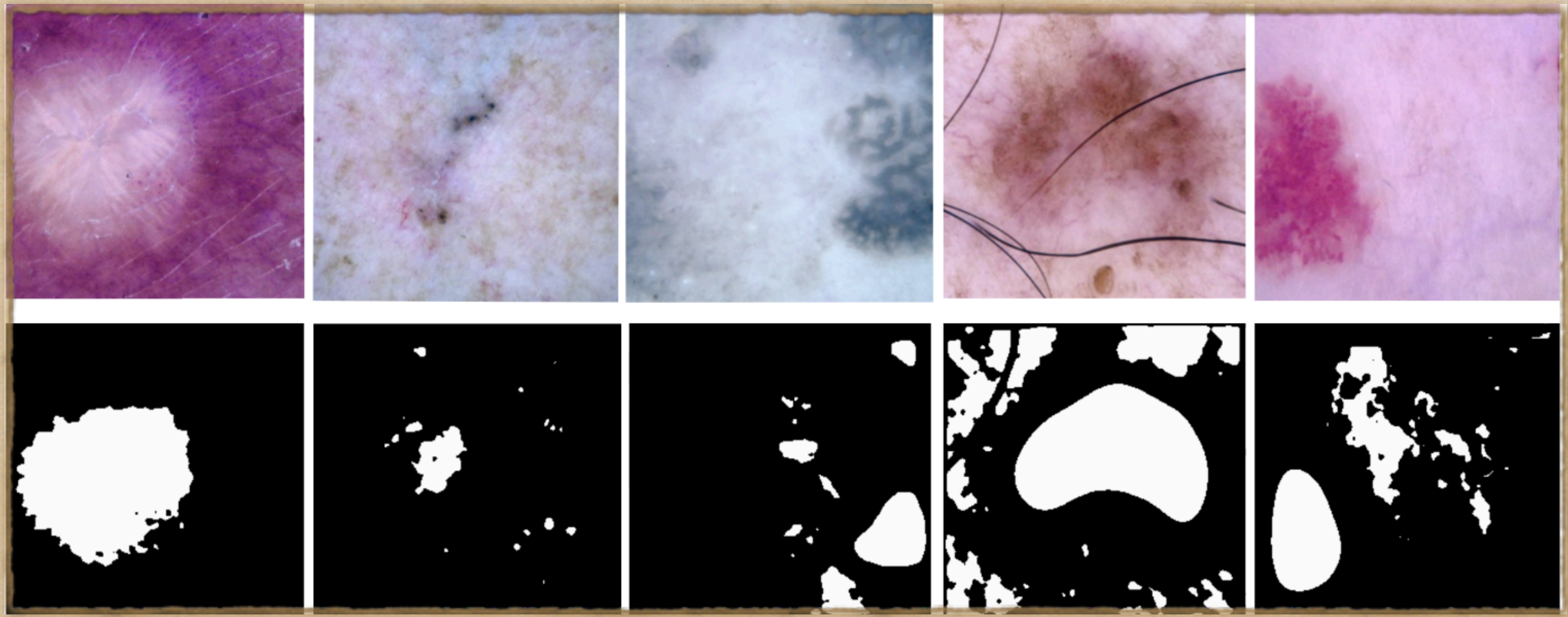




- Background
- Related Work
- Methodology
- Experiments
    - Setting up
    - Binary Classification
    - <u>Multi-class Classification</u>

ISIC2018 - “Fake” Lesion Segmentation

- Training U-Net on a small segmentation dataset (2594 images)
- Generating lesion segmentation of the classification training set (10015 images)
- Using the generated masks for attention regularization



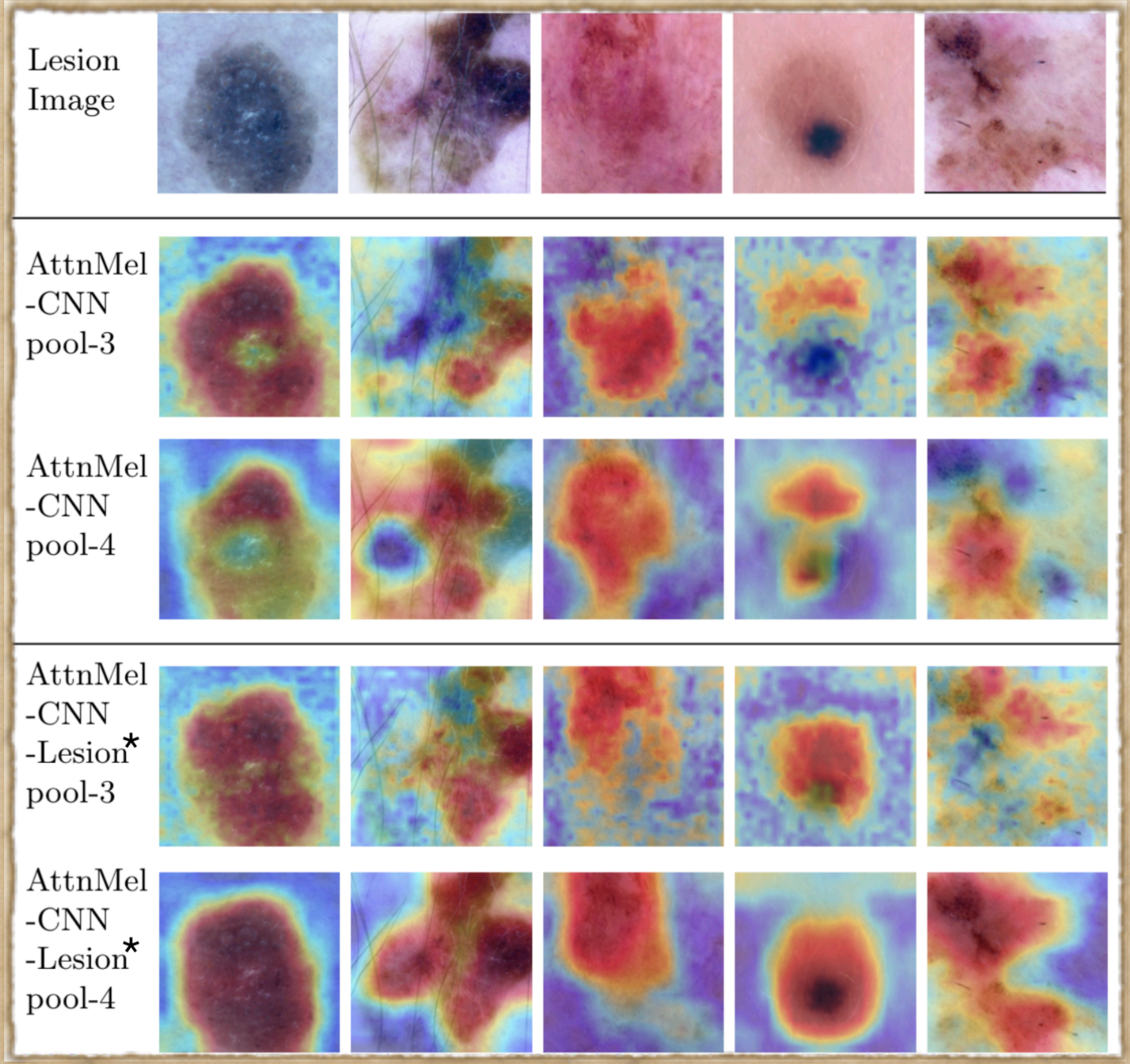
Results on ISIC2018

		MEL	NV	BCC	AKIEC	BKL	DF	VASC	AVG
#1	VGG-16	**0.829**	0.848	0.902	0.750	0.782	0.545	**1.0**	0.808
#2	VGG-16-GAP	0.811	0.870	0.863	0.813	0.845	0.545	**1.0**	0.821
#3	Mel-CNN	0.811	0.861	0.902	**0.906**	0.827	0.545	0.929	0.826
#4	AttnMel-CNN	0.784	**0.896**	**0.941**	0.813	0.818	**0.636**	**1.0**	0.841
#5	AttnMel-CNN-Lesion*	0.801	**0.896**	0.922	0.750	**0.873**	**0.727**	**1.0**	**0.853**

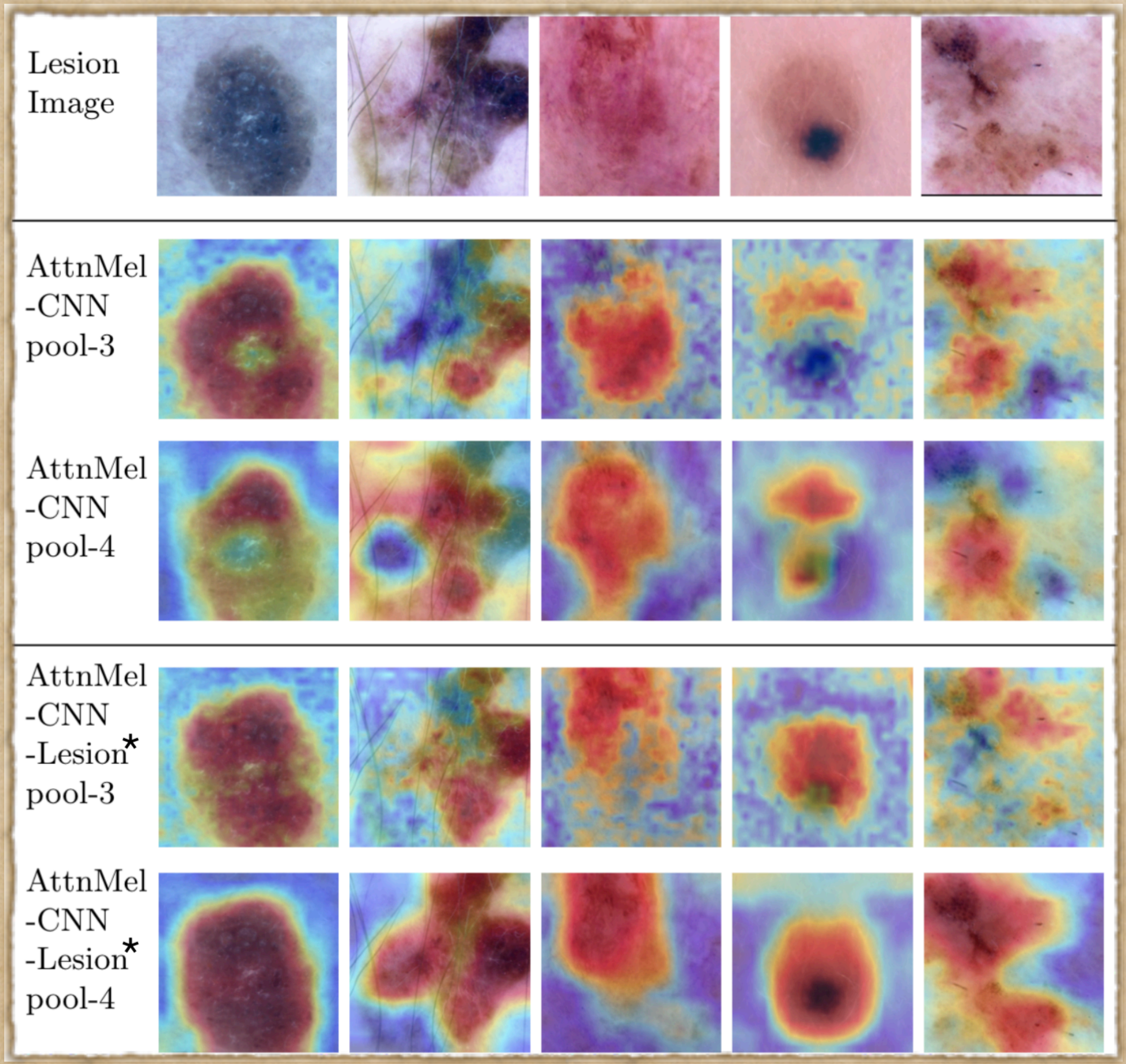
Results on ISIC2018

		MEL	NV	BCC	AKIEC	BKL	DF	VASC	AVG
#1	VGG-16	**0.829**	0.848	0.902	0.750	0.782	0.545	**1.0**	0.808
#2	VGG-16-GAP	0.811	0.870	0.863	0.813	0.845	0.545	**1.0**	0.821
#3	Mel-CNN	0.811	0.861	0.902	**0.906**	0.827	0.545	0.929	0.826
#4	AttnMel-CNN	0.784	**0.896**	**0.941**	0.813	0.818	**0.636**	**1.0**	0.841
#5	AttnMel-CNN-Lesion*	0.801	**0.896**	0.922	0.750	**0.873**	**0.727**	**1.0**	**0.853**

Results on ISIC2018



Results on ISIC2018

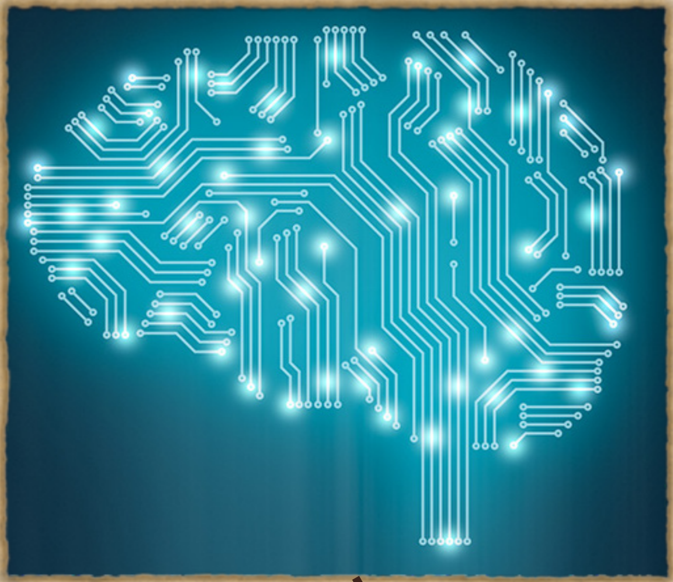


"Imperfect" attention regularization can also improve performance.

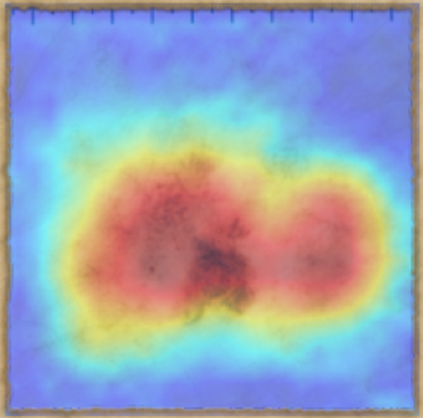
Conclusion

- Attention helps with skin cancer diagnosis;
- Attention regularization: a flexible and robust way of applying any types of pixel-level prior information;

Future Work: User Study



Machine



Human

Experts

References

- Zhuang et al. Skin lesion analysis towards melanoma detection using deep neural network ensemble. International Skin Imaging Collabo- ration (ISIC) Challenge on Skin Image Analysis for Melanoma Detection. MICCAI, 2018.
- Codella et al. Deep learning ensembles for melanoma recognition in dermoscopy images. IBM Journal of Research and Development, 61(4):1–15, 2017.
- Yu et al. Automated melanoma recognition in dermoscopy images via very deep residual networks. IEEE Transactions on Medical Imaging, 36(4):994–1004, 2017.
- Chen et al. A multi-task frame- work with feature passing module for skin lesion classification and segmentation. In IEEE International Symposium on Biomedical Imaging, pages 1126–1129, 2018.
- Molle et al. Visualizing convolutional neural networks to improve decision support for skin lesion classification. In MICCAI Workshop on Understanding and Interpreting Machine Learning in Medical Image Computing Applications, pages 115–123. Springer, 2018.
- Ge et al. Skin disease recognition using deep saliency features and multimodal learning of dermoscopy and clinical images. In International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI), pages 250–258. Springer, 2017.

Publication

- Yiqi Yan, Jeremy Kawahara, and Ghassan Hamarneh. Melanoma recognition via visual attention. In International Conference on Information Processing in Medical Imaging, Lecture Notes in Computer Science, vol 11492, pages 793–804, Springer, 2019. DOI https://doi.org/10.1007/978-3-030-20351-1_62

- https://github.com/SaoYan/IPMI2019-AttnMel

Q & A