Melanoma Detection Through Transfer Learning

By Anas Hamed

November 29th, 2019

Introduction

Skin Cancer

- Most common form of cancer in the U.S.
- Accounts for **40%** of all cases of cancer worldwide
- Three Types:
  - **Basal-cell**: not likely to spread
  - **Squamous-cell**: rarely results in any serious complications
  - **Melanoma**: most aggressive, most deadly, highest mortality rate
- Melanoma accounts for **75%** of cases that result in death
- **99%** survival rate if detected early and treated
- Only **7-14%** survival rate for advanced stages

Source: https://en.wikipedia.org/wiki/Skin_cancer

Melanoma Detection Through Images

Indicators

- **ABCDEF** mnemonic[3]:
  - **A**symmetry
  - **B**orders
  - **C**olor
  - **D**iameter
  - **E**volving
  - **F**unny-looking

Challenges

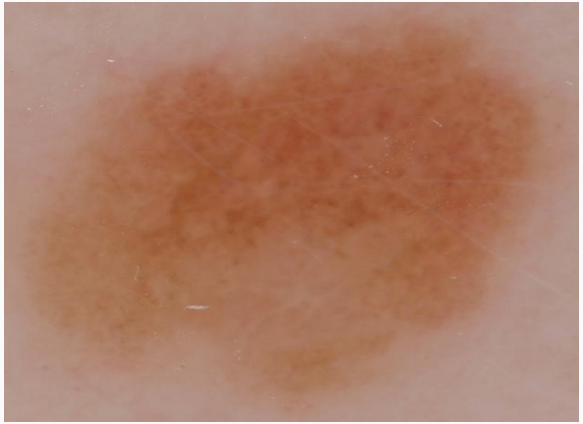
- A lot of malignant lesions closely resemble a benign nevus (mole)
- Not all characteristics can be captured in an image (e.g. firmness, elevation)

ISIC Challenge

- Skin Lesion Analysis Towards Melanoma Detection[2]
- Hosted by **International Skin Imaging Collaboration** (ISIC)
- Launched in 2018 and March 2019
- **159** participants (classification task)
- ~**24,000** dermoscopic images across 3 tasks:
    - Disease classification
    - Lesion segmentation
    - Attribute detection

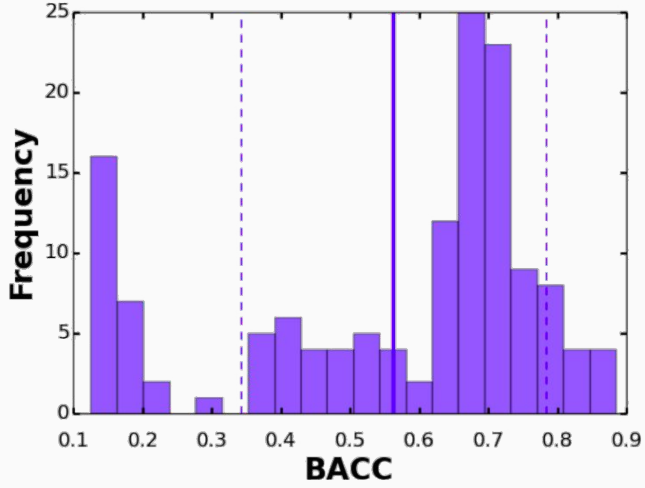
Dermoscopic Images

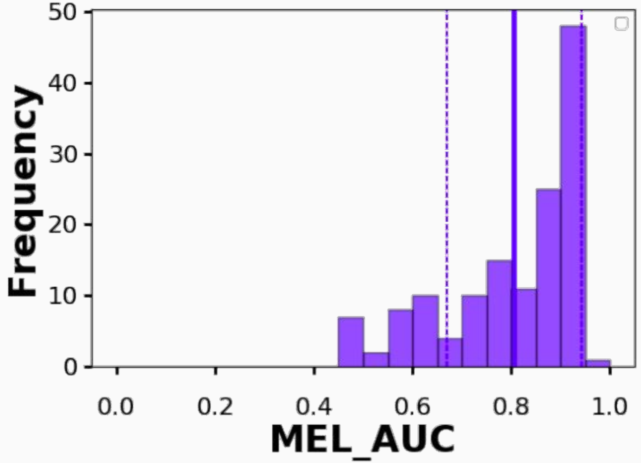
- Produces an illuminated, magnified image of the lesion of interest
- Provides for a more accurate classification compared to visual inspection by the naked eye



Evaluation[2]

- **Balanced Accuracy**
  - (TPR + TNR)/2
- **AUC**: Measures how well the model distinguishes between classes
  - An AUC of 1 means the model perfectly classifies each sample
  - An AUC of 0 means the model misclassifies all the samples





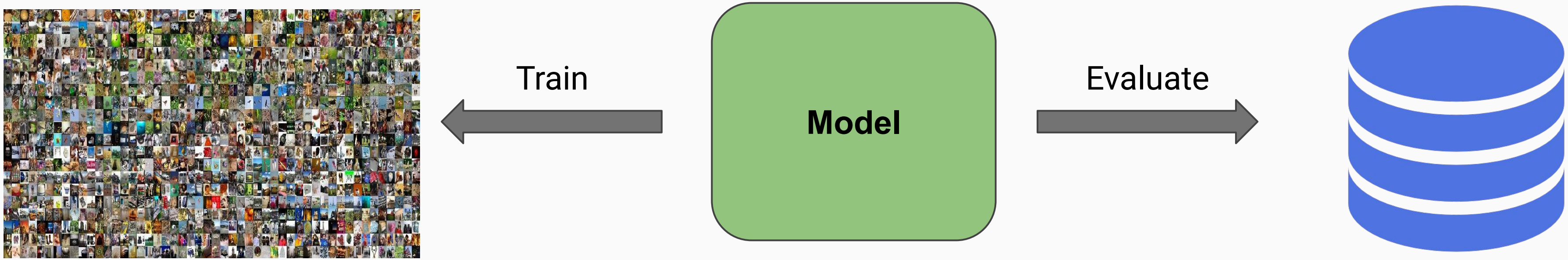
Experiments

Goals

- Demonstrate the effectiveness of state-of-the-art models and techniques
- Provide a baseline for future work
- Outline the challenges involved in melanoma detection through Deep Neural Networks

Transfer Learning

- A pretrained model is used for tasks involving feature extraction
- Two architectures (pretrained on ImageNet[4])
  - **ResNet-50**[5]
  - **ResNext-50**[6]



ImageNet (**1000** classes)

Dermoscopic Images

Dataset

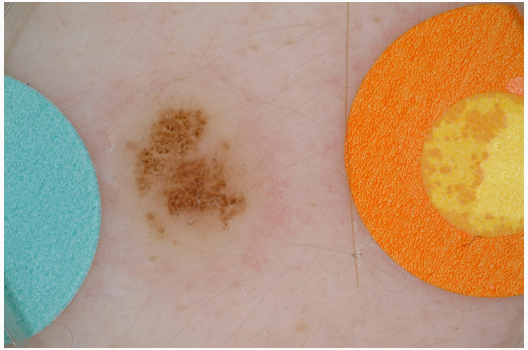
- ~**24,000** publicly available dermoscopic images + metadata
- Images were chosen that meet the following criteria:
  - Image has corresponding metadata file
  - Diagnosis is known
  - Segmentation mask is available

End result: ~**12,200** images

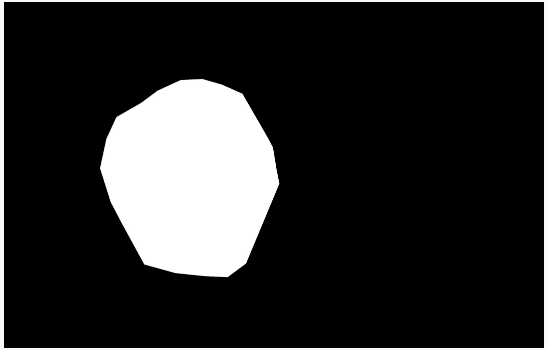
- 80% Training, 10% Validation and 10% Test

Dataset Preprocessing

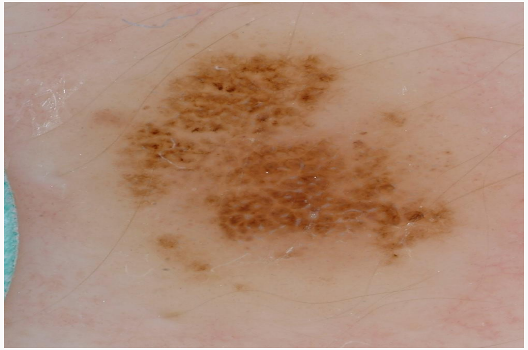
Original Image



Segmentation Mask



Result

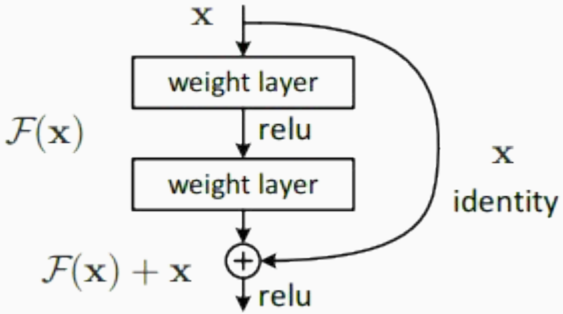


Dataset Preprocessing

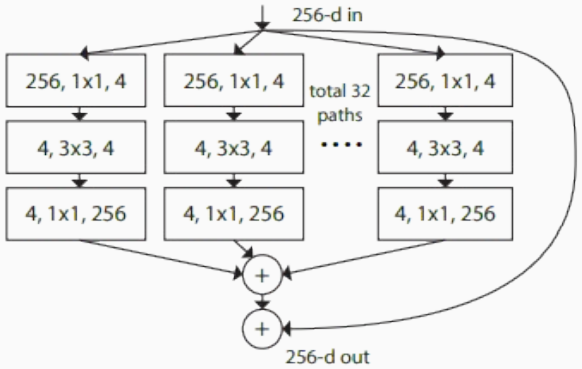
- Augmentation
  - Stochastic Horizontal Flip with **p = 0.5**
  - Stochastic Vertical Flip with **p = 0.5**
- Resizing
  - Original image size ~(1000-4000x800-2000 pixels)
  - Resized to **128x128**
- Normalization
  - For each pixel, subtract channel-wise mean and divide by standard deviation

Models

- **ResNet-50**[5]
  - Uses residual connections
  - Proven to be effective for classification tasks

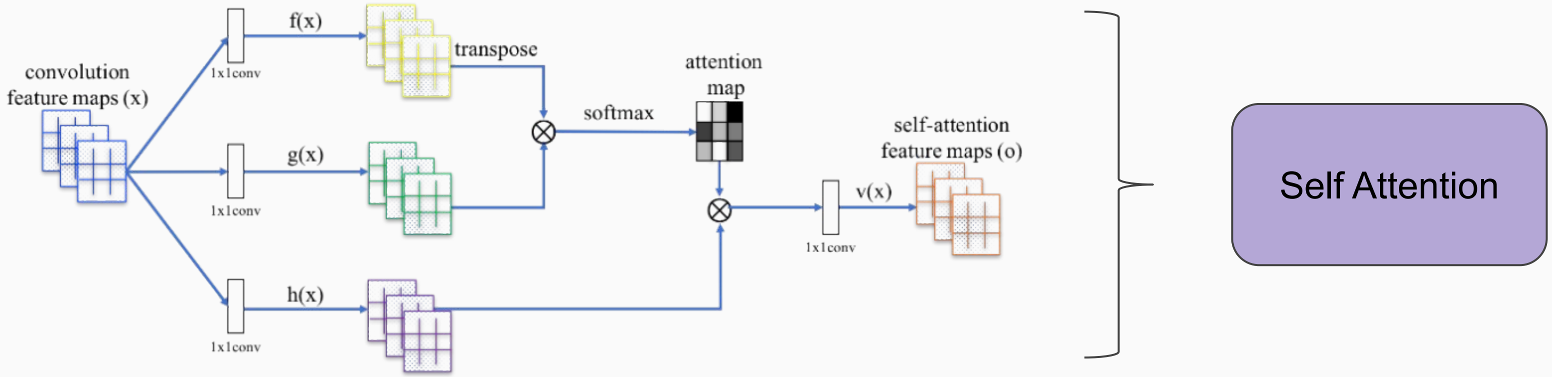


- **Resnext-50**[6]
  - Uses blocks of split-transform-merge operations
  - Outperforms some versions of the ResNet model with decreased complexity



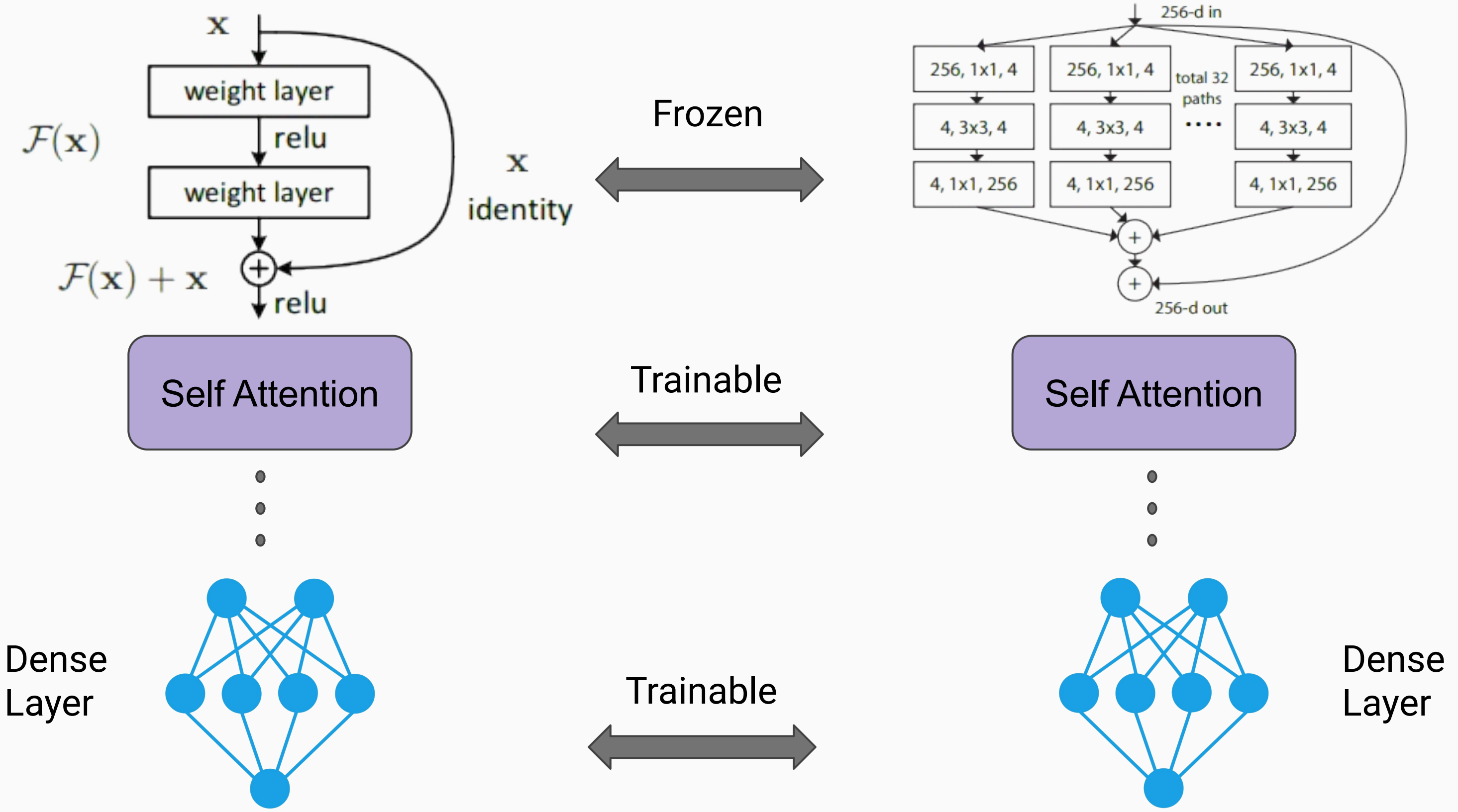
Self Attention[7]

- Uses attention maps to overcome limitations of convolution layers
- Places more importance on relevant locations in an image



Architecture

- ResNet-50
- ResNext-50

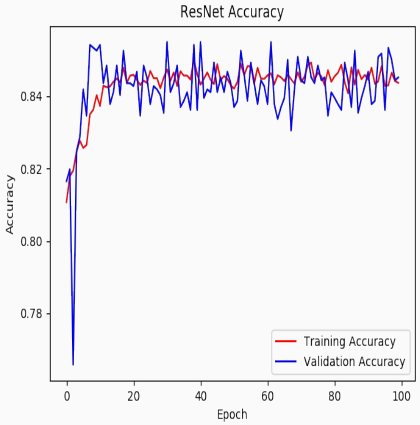


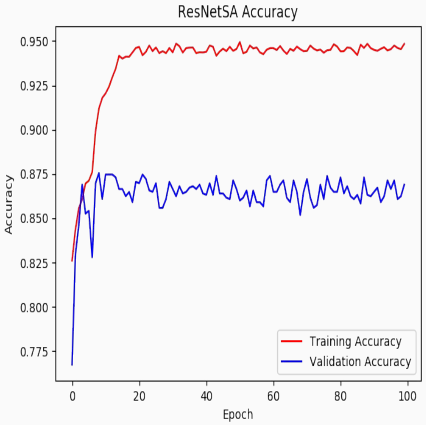
Training

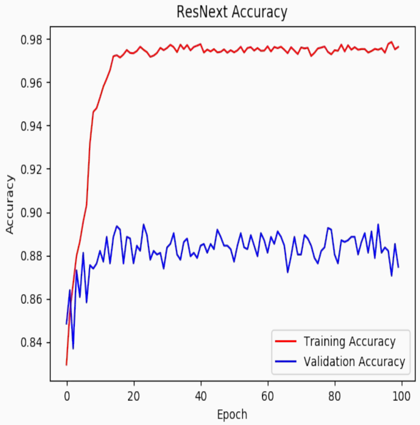
- **100** epochs
- Cross Entropy Loss[8]
- Adam Optimizer[9]
- **4** training cycles
    - ResNet with attention
    - ResNet without attention
    - ResNext with attention
    - Resnext without attention

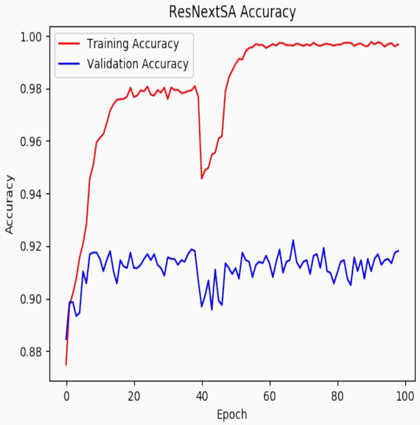
Results & Conclusion

Results


ResNet Accuracy
Training Accuracy
Validation Accuracy
Accuracy
Epoch
0.84
0.82
0.80
0.78
0
20
40
60
80
100


ResNetSA Accuracy
Training Accuracy
Validation Accuracy
Accuracy
Epoch
0.950
0.925
0.900
0.875
0.850
0.825
0.800
0.775
0
20
40
60
80
100


ResNext Accuracy
Training Accuracy
Validation Accuracy
Accuracy
Epoch
0.98
0.96
0.94
0.92
0.90
0.88
0.86
0.84
0
20
40
60
80
100


ResNextSA Accuracy
Training Accuracy
Validation Accuracy
Accuracy
Epoch
1.00
0.98
0.96
0.94
0.92
0.90
0.88
0
20
40
60
80
100

Results

	Test Accuracy (%)	Balanced Accuracy (%)	Precision	Recall	F1 Score	Training Time (100 epochs)
ResNet	83.86	75.91	0.63	0.41	0.497	14.7 hrs
ResNetSA	86.56	78.79	0.65	0.44	0.527	14.8 hrs
ResNext	87.46	79.12	0.64	0.54	0.592	14.7 hrs
ResNextSA	**91.04**	**80.33**	**0.67**	**0.56**	**0.612**	**14.9 hrs**

Further Improvements

- Adding more attention layers
- Ensembling
- Using deeper models (e.g. ResNext-101)
- Experimenting with different models
- Increasing size of dataset
- Incorporating sample metadata
- Countering imbalance between class samples

References

[1] Project Code: https://github.com/AnasHamed73/676-deep-learning-project2

[2] ISIC Challenge:https://arxiv.org/pdf/1902.03368.pdf

[3] Melanoma: https://en.wikipedia.org/wiki/Melanoma

[4] ImageNet: http://www.image-net.org/

[5] ResNet: https://arxiv.org/pdf/1512.03385.pdf

[6] ResNext: https://arxiv.org/pdf/1611.05431.pdf

[7] Self Attention: https://arxiv.org/pdf/1706.03762.pdf

[8] Cross Entropy Loss: https://en.wikipedia.org/wiki/Cross_entropy

[9] Adam Optimizer: https://arxiv.org/pdf/1412.6980.pdf