Cancer Diagnosis with Accuracy and Explainability

Group 24: Chuqiao Yuan, Queena Zhou, Rong Liu, Feiyu Chen

Introduction

GOAL

Explainable classifier for cancer diagnosis

Accuracy

Correct Diagnosis

Explainability

Regions of interest (ROI) localization

Methodology

ResNet

Residual learning with deep structure

Binary classification and saliency map

UNet

Semantic segmentation

UNet with ResNet Backbone

UNet with shortcut connections

Stacking

Support Vector Machine

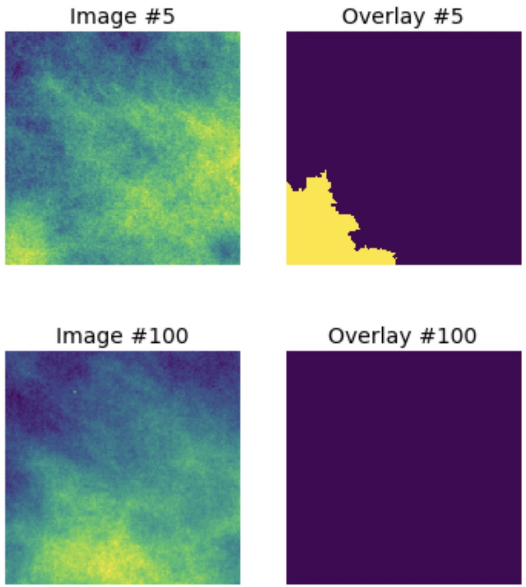
Data

Dataset: CBIS-DDSM (for breast cancer)

- 1318 training cases, 378 testing cases
- Original X-ray image with overlay annotating ROI

Preprocessing

- Shape of image is large and varied
- Cropped to 256*256 patches
- Normalized to [0,1]
- Filtered to maintain only patches with breast organization
- 19702 training images, 5835 testing images



Metrics

Accuracy

Diagnosis Accuracy

$$\frac{\text{Number of images classified correctly}}{\text{Total number of images}}$$

Explainability

Pixel Accuracy

$$\frac{\text{Number of pixels classified correctly}}{\text{Total number of pixels}}$$

Intersection over Union

$$\frac{\text{Intersection of problem pixels}}{\text{Union of problem pixels}}$$

Baseline

Model	Diagnosis Acc.	Pixel Acc.	IoU
Predicting all pixels to be normal	0.700	0.922	0.000
Predicting all pixels to be abnormal	0.300	0.078	0.078

Strongly Imbalanced:
Normal pixels outnumber abnormal ones

ResNet

Overview

Motivation

Powerful image classification model

Principles

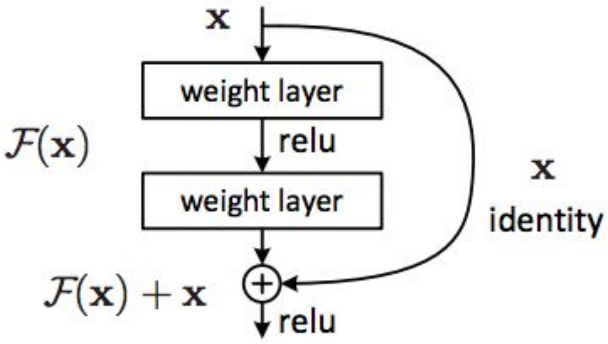
"Identity shortcut connection" + fit a residual mapping

Expectation

High diagnosis classification accuracy

Suspicion

Limited help in localization



(He, Zhang, Ren, Sun.. 'Deep Residual Learning for Image Recognition'. 2015.)

ResNet

Saliency Map

1. **ResNet Model**: Select the model without overfitting
2. **Last Layer**: Switch the softmax activation for linear
3. **Saliency Map**: Produce saliency maps for images in test data
4. **Normalization**: Normalize the values in each saliency map
5. **Classify Each Pixel**: Pixels with values more than 0.5 to be classified as abnormal, otherwise normal

UNet

Overview

Motivation

Directly output ROI annotations

Principle

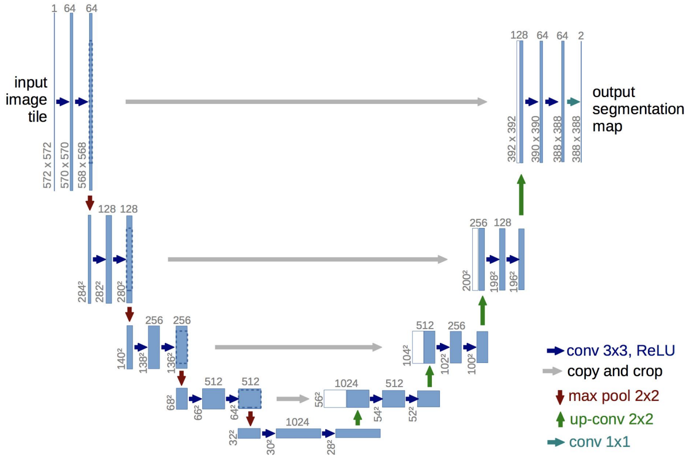
Upsampling layers + Copying over feature maps

Expectation

High IoU measure

Usage for Diagnosis

Whether the generated annotation contains any problem pixel



(Ronneberger, Fischer, Brox. 'U-Net: Convolutional Networks for Biomedical Image Segmentation'. 2015.)

UNet
Tuning

Weight in Loss	# Filters	# Blocks	Diagnosis Acc.	Pixel Acc.	IoU
Balanced	64x	9	0.292	0.071	0.071
Balanced	32x	9	0.622	0.858	0.234
Reduced by 1/3	32x	9	0.592	0.857	0.236
Reduced by 1/2	**32x**	**9**	**0.680**	**0.906**	**0.267**
Reduced by 2/3	32x	9	0.708	0.929	0.000
Reduced by 1/2	32x	11	0.708	0.929	0.000

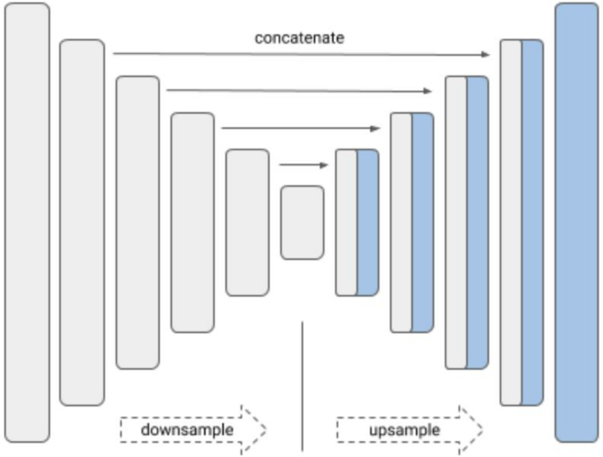
UNet with ResNet Backbone

Motivation

- ResNet: add **short skips** to avoid vanishing gradient
- UNet: add **long skips** to recover spatial information
- We can have a network that has **both long and short skips**

UNet with ResNet Backbone

Architecture

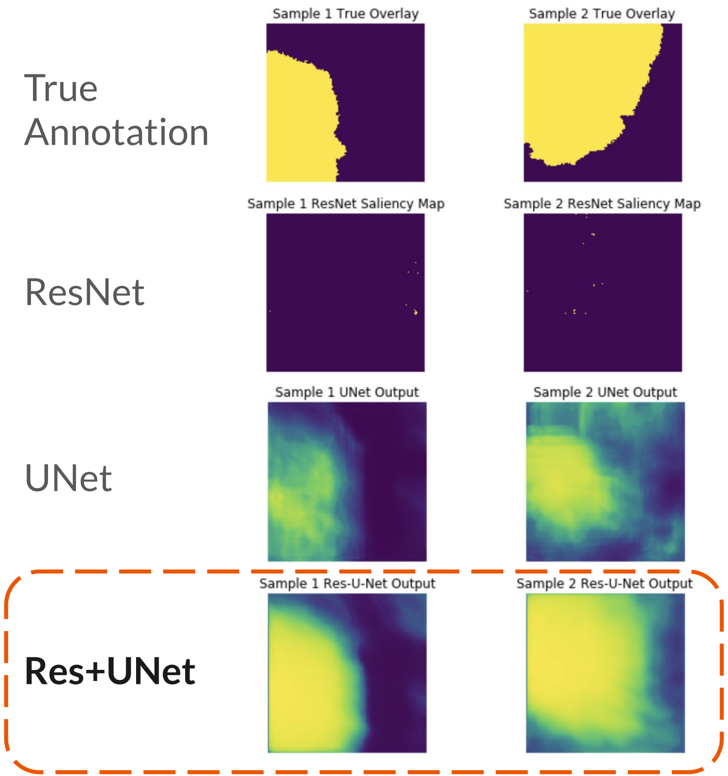


https://github.com/qubvel/segmentation_models

- Contracting path and expanding path are both ResNets
- We tried ResNet with **various depth**: ResNet-34/50/101
- **ResNet-34** performs the best

Comparing the Models
Generated ROI Annotation

- Both UNet and Res+UNet are able to identify problem pixels
- Res+UNet is more confident in locating problem pixels



Comparing the Models

Metrics

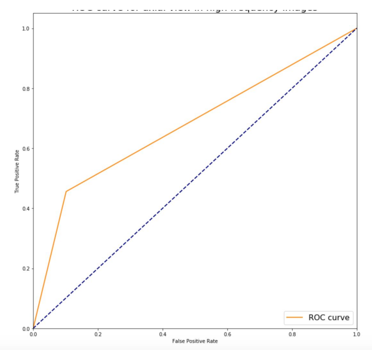
Model	Diagnosis Accuracy	Pixel-by-Pixel Accuracy	Intersection over Union
ResNet	**0.769**	0.881	0.007
UNet	0.708	**0.911**	0.269
UNet with ResNet Backbone	0.705	0.898	**0.330**

There is a **tradeoff** between accuracy and explainability

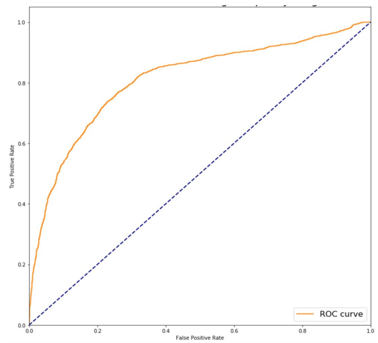
Stacking
Real Life Medical Application

- ResNet has a high false negative rate on classification, which leads to a large amount of unalarmed tumors
- UNet has a high false positive rate on classification, which leads to medical resource waste
- Stacking of the models can achieve a higher accuracy (0.78) and a higher AUC (0.81)
- The final deliverable package will give out both prediction and heat map, available on https://github.com/YuanChuqiao/Diagnosis

ROC for ResNet



ROC for Stacking Model



Conclusion and Takeaways

1. Each model has its pros and cons.
2. Stacking models reduces false negative rate.
3. We recommend doctors to use the stacking model.
4. Future work includes:
   a. Include data augmentation
   b. Try larger input image size
   c. Try ResNet100 and ResNeXt50