CLASSIFYING BIOLOGICAL IMAGES USING PRE-TRAINED CNNs

TEAM

Nikhil Makkar, Hasnain Bilgrami,
Hadi Mansourifar, Aditi Singh, Tanay Shah

1 Problem

Object Classification in Biological Domain

- Difficult to use a CNN on a small dataset
- Large number of samples required for greater accuracy
- Annotating biological images needs expertise
- Large labelled dataset not available

2 Dataset Description

Breast Cancer Histopathological Database

- Benign:
  - Adenosis
  - Fibroadenoma
  - Phyllodes tumor
  - Tubular Adenoma

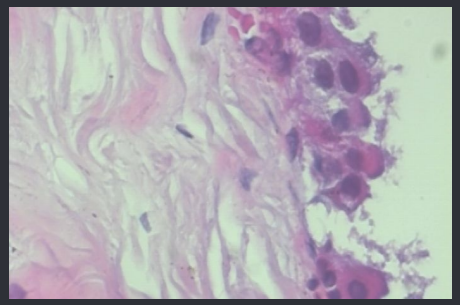
- Malignant:
  - Carcinoma
  - Lobular Carcinoma
  - Mucinous Carcinoma
  - Papillary Carcinoma

Breast Cancer Histopathological Database

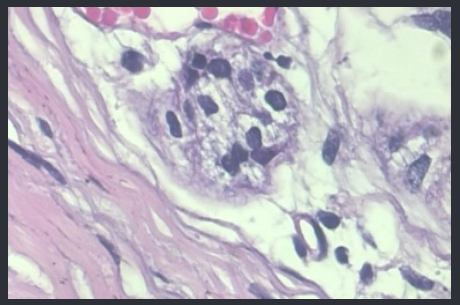
- Data from 82 patients
- RGB images, size 700x460

Magnification	Benign	Malignant	Total
40X	652	1,370	1,995
100X	644	1,437	2,081
200X	623	1,390	2,013
400X	588	1,232	1,820
Total Number of Images	**2,480**	**5,429**	**7,909**

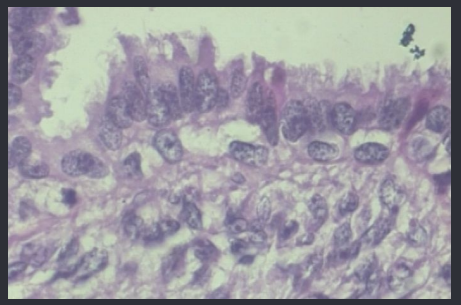
Samples at 400X magnification



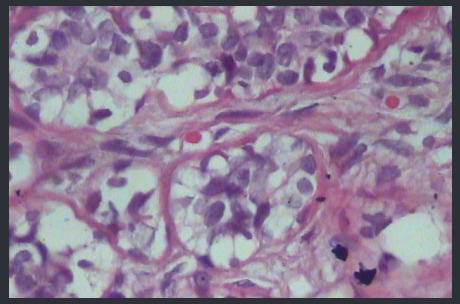
Adenosis



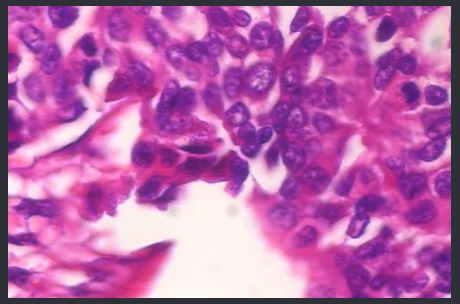
Fibroadenoma



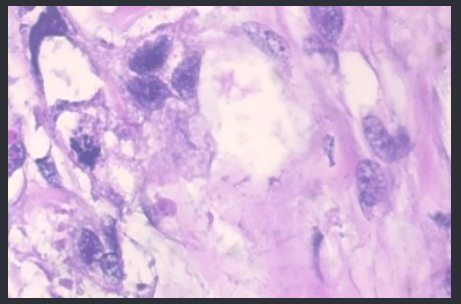
Phyllodes Tumor



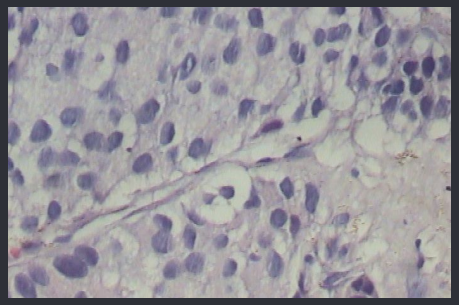
Tubular Adenoma



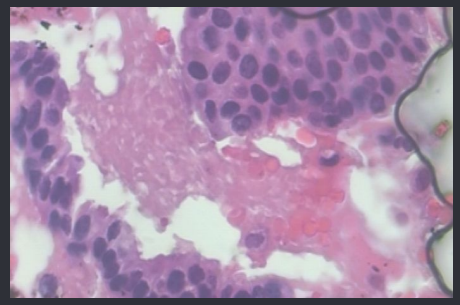
Carcinoma



Lobular Carcinoma



Mucinous Carcinoma



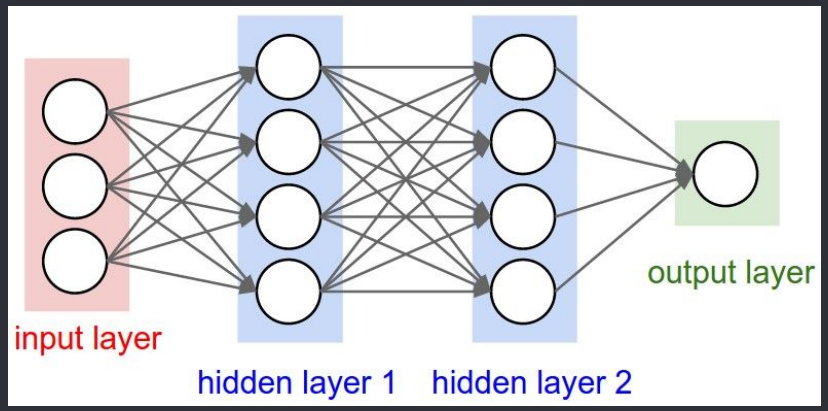
Papillary Carcinoma

3 Convolutional Neural Networks


LET ME SPRINKLE
SOME
DEEP LEARNING MAGIC
memegenerator.net

Convolutional Neural Network

- Similar to Neural Networks



- Encode images into meaningful features

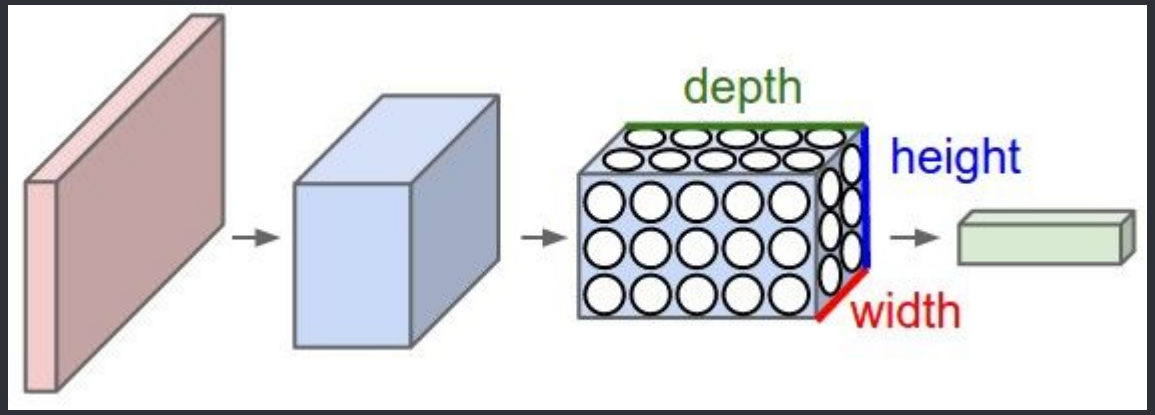
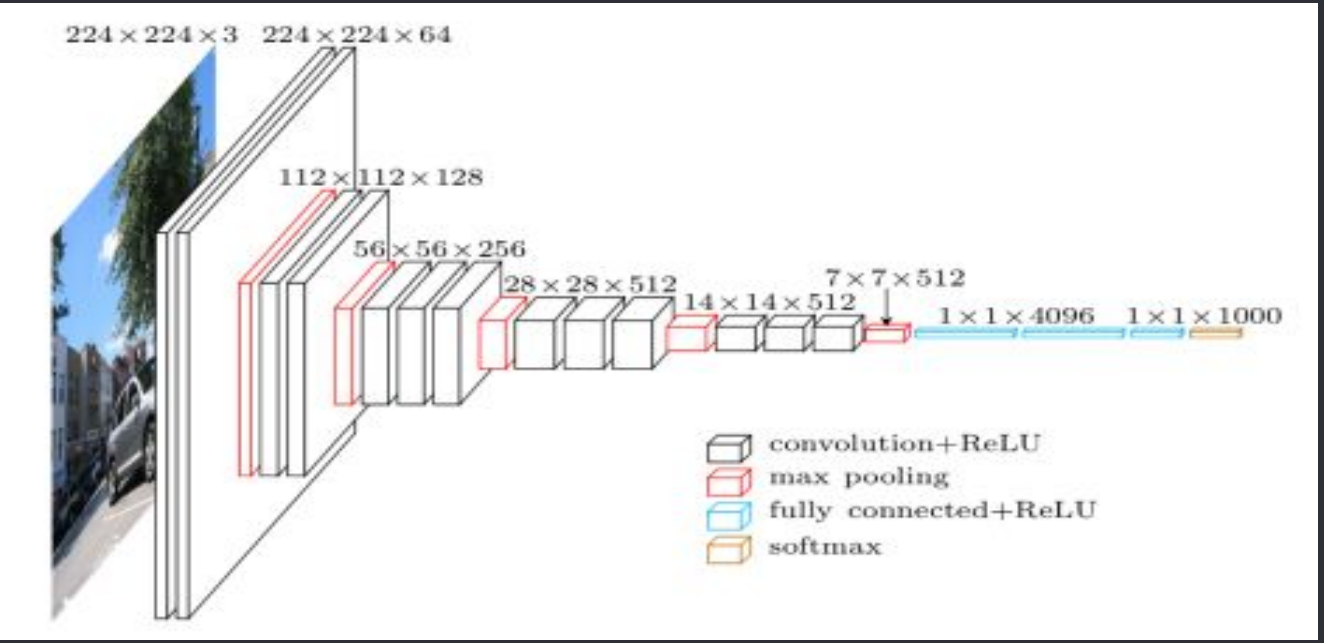


Image references: Convolutional Neural Networks for Visual Recognition

VGG-16 Network

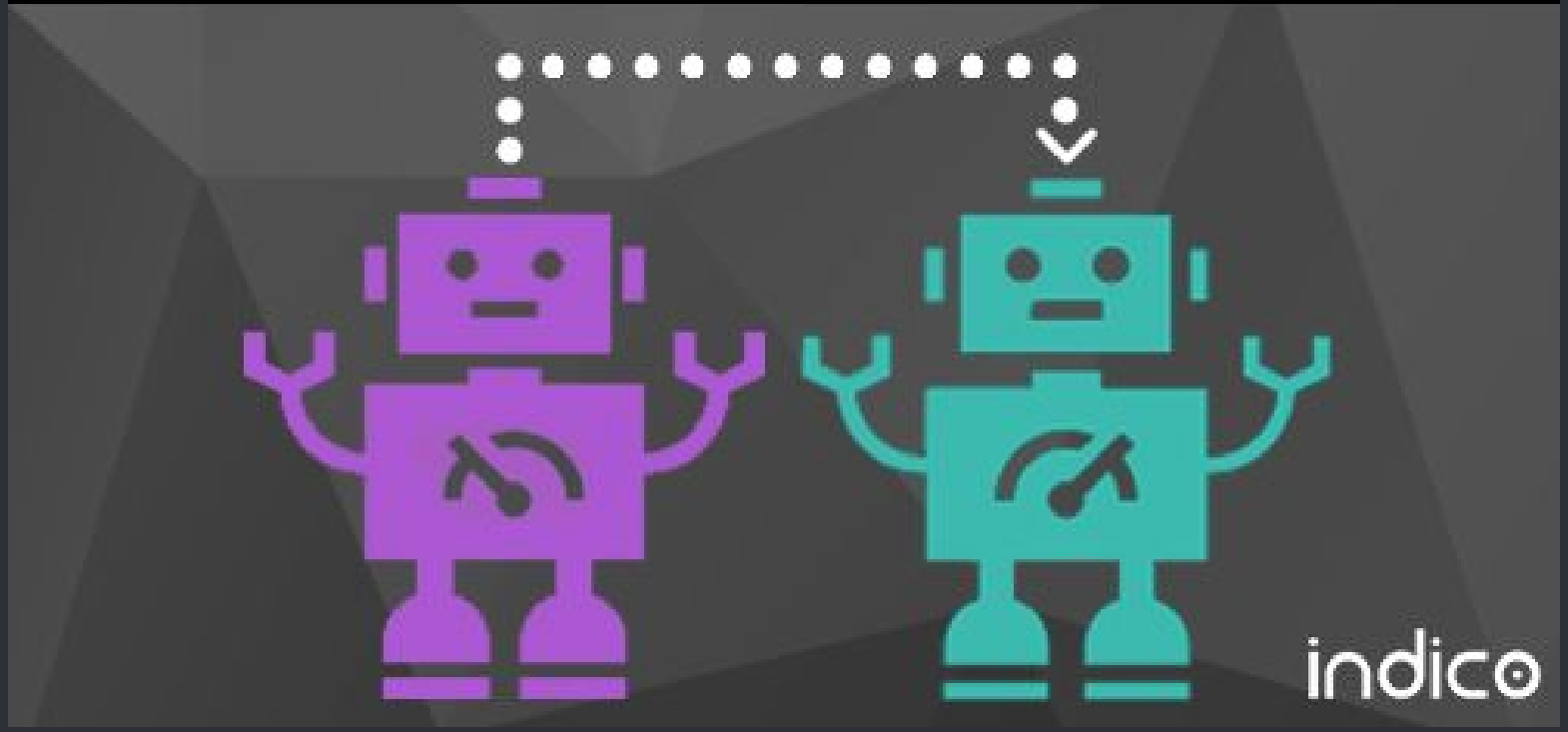


- Depth : critical component in performance
- Components: Conv layers, FC layers, ReLU, Pooling
- 140M parameters

Image reference: Heuritech Blog

4 Transfer Learning

transfer learning



ConvNet as Feature Extractor

- Use VGG-16 trained on Imagenet (1.2 M images, 1000 classes)
- Remove the last fully-connected layer
- Use as feature extractor, 4096 D vector for every image

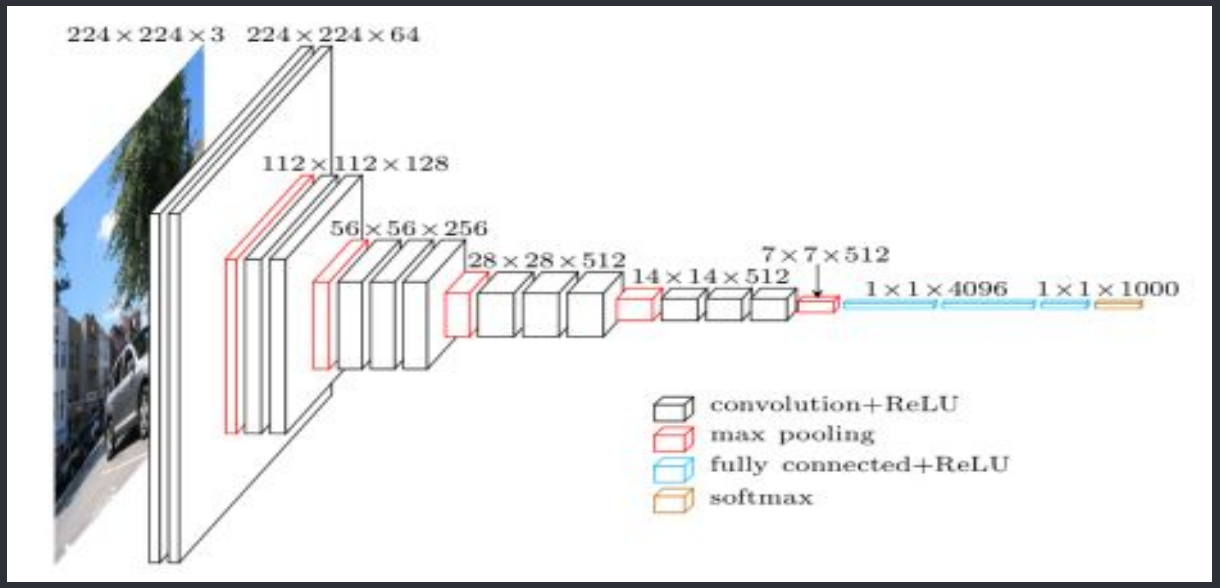
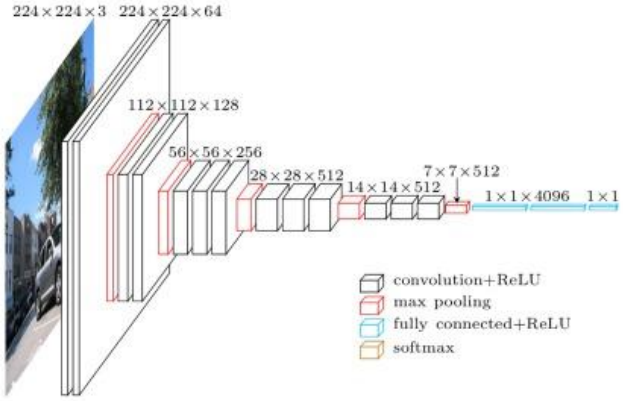


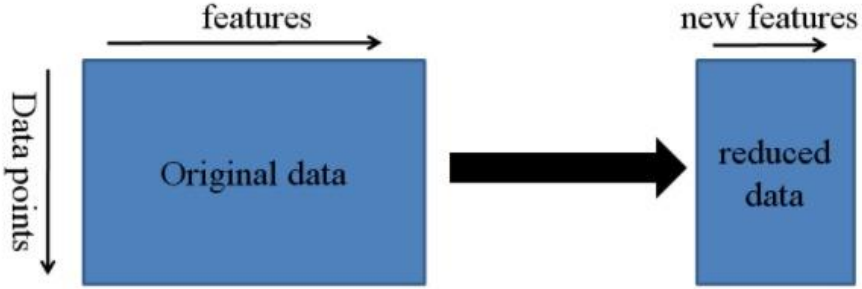
Image reference: Heuritech Blog

5 Approach

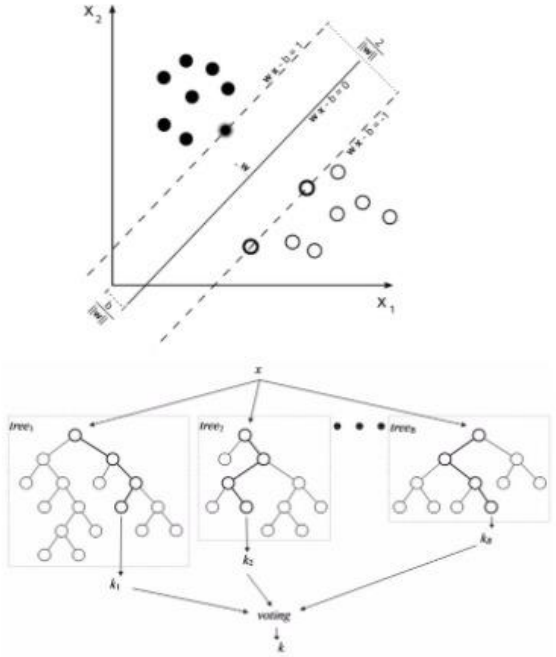
Step 1: Import database



Step 2: Getting the features from VGG-16



Step 3: Class imbalance



Step 4: SVM and Random Forest

6 Results

BreakHis Classification

- Dataset is divided according to the magnifications i.e. 40X, 100X, 200X, 400X
- 2 datasets:
  - 2 classes: Benign and Malignant
  - 8 classes: Subtypes of Benign and Malignant
- Classification is done using C-SVC linear, polynomial SVM and Random Forest

BreakHis Classification

2 classes (average accuracy)

	40X	100X	200X	400X
C-SVC Linear SVM	89%	89%	88%	88%
C-SVC Polynomial SVM	88%	90%	89%	85%
Random Forest	89.18%	88%	87.74%	80%

BreakHis Classification

8 classes (average accuracy)

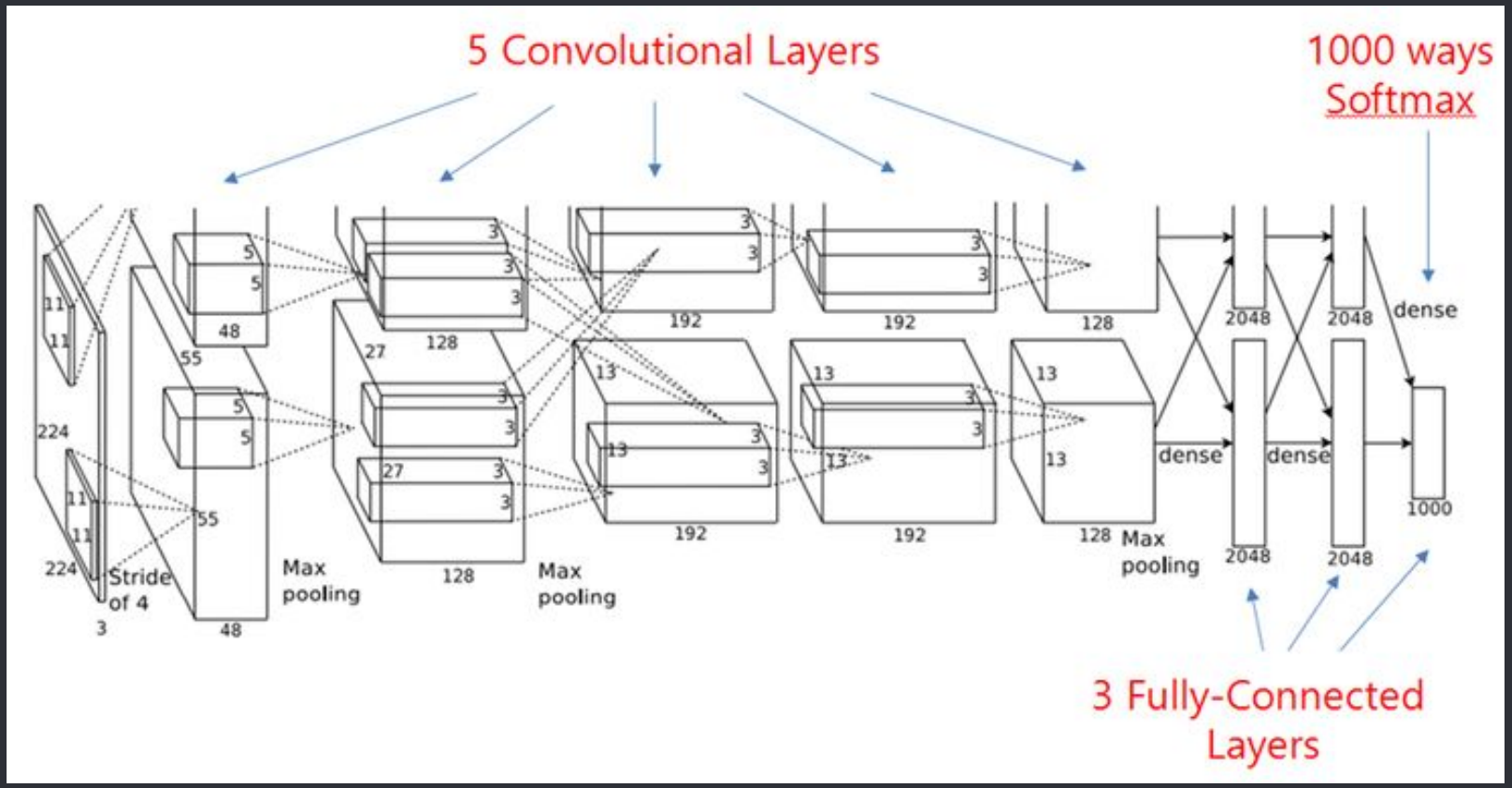
	40X	100X	200X	400X
C-SVC Linear SVM	55.97%	51.59%	55.93%	53.06%
C-SVC Polynomial SVM	53.70%	55,12%	54.78%	53.9%
Random Forest	48.50%	49%	49.81%	48.16%

7 Conclusion

- Pre-training can be used in case of small dataset
- Pre-trained network on natural images, can give significant result in images from an entire different domain
- Performance also depends on complexity of classification task

8 Ongoing Work

AlexNet



AlexNet: ~60M parameters

REFERENCES

- Breast Cancer Histopathological Database http://web.inf.ufpr.br/vri/breast-cancer-database
- F. A. Spanhol, L. S. Oliveira, C. Petitjean and L. Heutte, "Breast cancer histopathological image classification using Convolutional Neural Networks," 2016 International Joint Conference on Neural Networks (IJCNN), Vancouver, BC, 2016, pp. 2560-2567.
- Aravindh Mahendran,Andrea Vedaldi, “Visualizing Deep Convolutional Neural Networks Using Natural Pre-Images”
- Nima Tajbakhsh et. al., “Convolutional Neural Networks for Medical Image Analysis: Full Training or Fine Tuning?”
- http://www.deeplearningmodel.net
- https://github.com/tensorflow/models/tree/master/slim

Thanks!

ANY QUESTIONS?