COMP 448/548: Medical Image Analysis

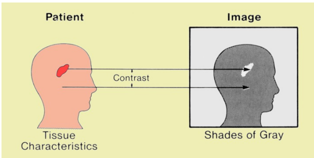
Review of the semester

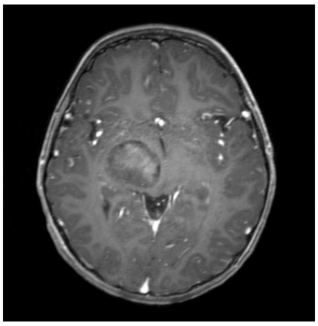
Çiğdem Gündüz Demir
cgunduz@ku.edu.tr

1

Medical imaging

- Process of converting tissue characteristics into a visual image
- For a tissue to be seen by a human or a computer, the imaging modality should have sufficient
  - **Contrast resolution:** ability to distinguish intensity differences between this tissue and its surrounding tissues
  - **Spatial resolution:** ability to visualize small enough tissues
- Different medical imaging modalities reveal different characteristics of the human body
  - No medical image (or no medical imaging modality) reveals everything
  - Visibility of specific features depends on the characteristics of the imaging modality and the manner it is operated
  - Contrast resolution and spatial resolution vary with the imaging modality and the manner it is operated

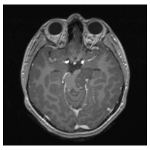




2

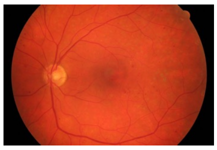
Medical images

- Today, imaging systems are extensively used in medicine and biology research
- Their primary use is to visualize human body at different levels
- Images are visually analyzed by clinicians/biologists to make decisions
- These analyses rely on visual interpretations/judgments



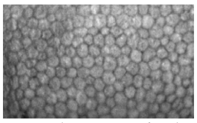
MR scan showing tumor in brain

SPECT-CT bone scan showing joint inflammation

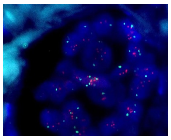


Fundus photography for diabetic retinopathy diagnosis

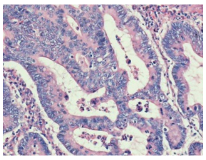
Skin photography for melanoma screening



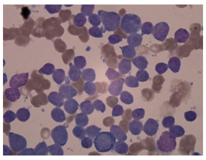
Corneal in vivo confocal microscope photography of endothelial cells for diagnosis



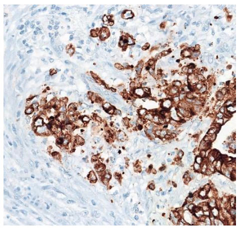
FISH to determine HER2 protein status in breast cancer



Micrograph of HE stained biopsy for colon adenocarcinoma grading



Micrograph of peripheral blood smear for leukemia diagnosis



Micrograph of IHC stained biopsy to localize tumor buddings, a prognostic indicator for cancer

3

Why do we need computers?

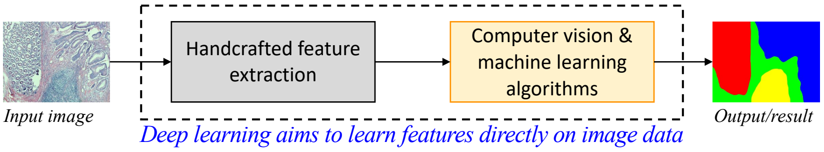
- We need computer help for medical image analysis since
  - There are *too many images* to be analyzed
  - Some tasks are easy for human observers but *time-consuming*
  - Some tasks are not that easy and heavily rely on human interpretations, which may lead to *intra- and inter-observer variability*
  - Analyses mostly rely on qualitative visual analyses but quantitative metrics are useful
- Computational tools facilitate rapid analyses with better reproducibility
- **The goal is to go from medical images to understanding**

4

Computational analysis tools

- For computers to understand medical images, they must
  - Represent an image with mathematical features and
  - Use these features in the design of their algorithms

- Most commonly studied algorithms are for ***segmentation and classification***



5

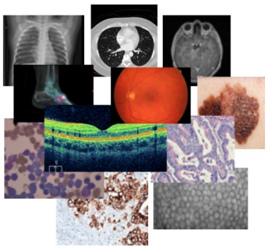
What should I have learned by now?

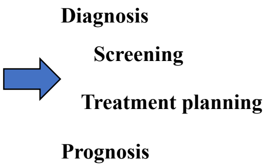
1. **General notion on medical imaging modalities**
2. Preliminaries for an algorithm design
3. Common challenges in medical image analysis
4. Traditional methods for image representation (feature extraction)
5. Medical image classification
6. Medical image segmentation
7. General notion on other commonly studied problems
   - Such as image synthesis, image denoising, image registration, super-resolution imaging, and multiple instance learning

6

Medical imaging is broad

- There are many medical imaging modalities used in
  - Pathology
  - Radiology
  - Nuclear medicine
  - Ophthalmology
  - Dermatology
  - …





- Although image analysis algorithms/systems share many common aspects, you need to know the domain and the imaging modality to more wisely design an algorithm/system

7

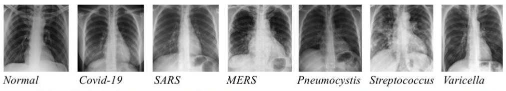
What should I have learned by now?

1. General notion on medical imaging modalities
2. **Preliminaries for an algorithm design**
3. Common challenges in medical image analysis
4. Traditional methods for image representation (feature extraction)
5. Medical image classification
6. Medical image segmentation
7. General notion on other commonly studied problems
   - Such as image synthesis, image denoising, image registration, super-resolution imaging, and multiple instance learning

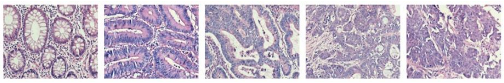
8

Input and output

- Inputs: 2D image, 3D volume, video
- Outputs: Classification or regression
  - Single output for an entire image/volume
  - A map of outputs for image pixels



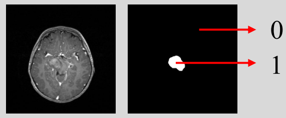
Chest X-ray classification



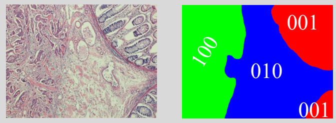
Tissue image classification

In instance segmentation, the annotation map may contain different labels for each instance → still *binary classification* if all instances belong to the same class

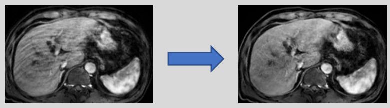
Foreground / background segmentation is very common → *binary classification*



More than two segmentation labels → *multiclass classification*



Artifact reduction in MR images



Input: Raw MR image whose pixels are in [0, 255]

Output: Artifact-reduced MR image whose estimated pixels are also in [0, 255]

9

Design pipeline

Problem definition and dataset preparation
- Define a problem
- Collect data
- Annotate data

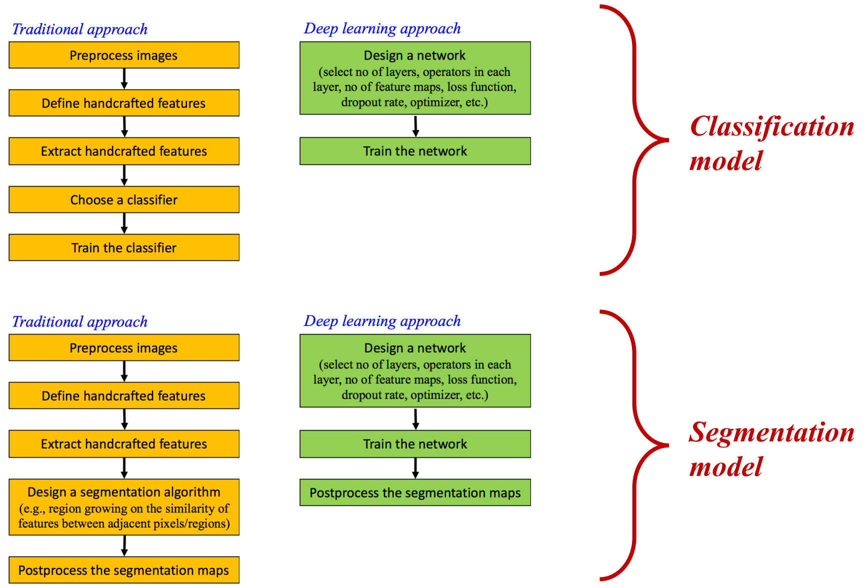
Build an image analysis model
- Design an algorithm
- Select parameters (if any)
- Train the model (if required)

Evaluate the model
- Evaluate the model performance *visually and quantitatively*
- Analysis *comparison with the existing approaches, ablation studies, and parameter analysis*

- Each step has its own challenges to overcome
- Model evaluation
  - Be aware of the bias and variance in performance metrics
  - Follow "proper" steps for model evaluation
    - Parameter selection
    - Comparative and ablation studies
    - Parameter analysis
    - Statistical tests

10

How to build an image analysis model?



11

How to quantitatively evaluate classification results?

Accuracy

- Percentage of correctly classified samples
- We may also want to consider class-based accuracies, especially when there is an unbalanced distribution among classes

Confusion matrix

Predicted class (columns), *True class* (rows)

	C_1	C_2	…	C_C
C_1				
C_2				
…				
C_C				

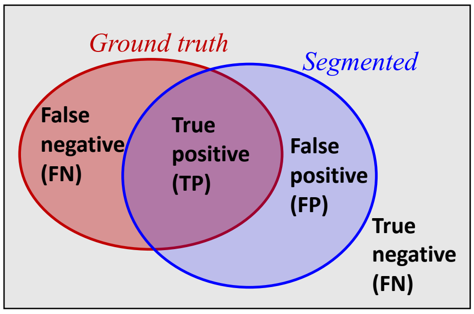
Do not use the same set of (training) samples both for learning a model and its evaluation!!!

- If available, use an independent test set
- If not, create multiple independent training and test sets by partitioning samples many times
  - Bootstrapping, k-fold cross-validation, leave-one-out

12

How to quantitatively evaluate segmentation results?

- Pixel-level evaluation



$$\text{Sensitivity} = \frac{TP}{TP + FN} \qquad \text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN}$$

$$\text{Specificity} = \frac{TN}{TN + FP} \qquad \text{Dice coeff} = \frac{2\,TP}{2\,TP + FP + FN}$$

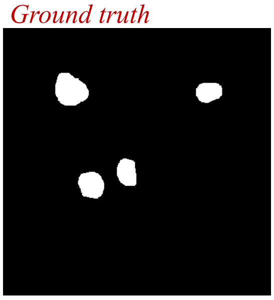
$$\text{Recall} = \frac{TP}{TP + FN} \qquad \text{F-score} = 2\,\frac{\text{Precision . Recall}}{\text{Precision + Recall}}$$

$$\text{Precision} = \frac{TP}{TP + FP}$$

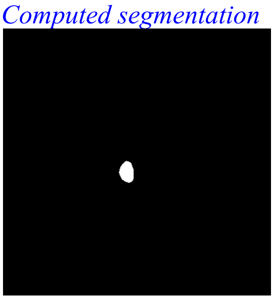
How to quantitatively evaluate segmentation results?

- Sometimes, pixel-level evaluation may be misleading

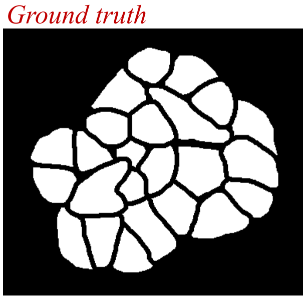
Ground truth



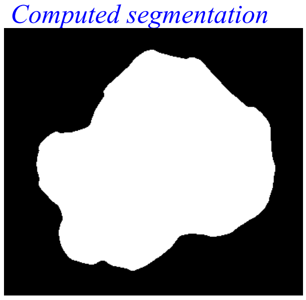
Computed segmentation



Ground truth



Computed segmentation



Sensitivity = 0.1071
Specificity = 1.0000
Accuracy = 0.9719
Dice coeff = 0.1935

Recall = 0.1071
Precision = 1.0000
F-score = 0.1935

Sensitivity = 0.9418
Specificity = 0.8564
Accuracy = 0.8883
Dice coeff = 0.8633

Recall = 0.9418
Precision = 0.7969
F-score = 0.8633

How to quantitatively evaluate segmentation results?

- Object-level evaluation
  - Necessary for instance segmentation tasks
  - Need to match segmented instances with the ground truth objects

***Object-level F-score* is to assess what percentage of instances are correctly detected

***Object-level Dice index* is to assess how accurately the pixels of the segmented instances overlap with those of their matching (maximally overlapping) ground truth objects

***Intersection over union (IoU)* is also to assess how accurately the pixels of the segmented instances and their matching ground truth objects overlap

***Object-level Hausdorff distance* is to assess the shape similarity between the segmented instances and their matching ground truth objects

15

What should I have learned by now?

1. General notion on medical imaging modalities
2. Preliminaries for an algorithm design
3. **Common challenges in medical image analysis**
4. Traditional methods for image representation (feature extraction)
5. Medical image classification
6. Medical image segmentation
7. General notion on other commonly studied problems
   - Such as image synthesis, image denoising, image registration, super-resolution imaging, and multiple instance learning

16

Common challenges

- Defining a "*meaningful*" problem is typically hard and requires familiarity with the associated medical domain
- Data collection takes effort and time
- Annotation is very challenging
  - It needs medical expertise
  - There might be inconsistencies in annotations
    - Sometimes there is no consensus among annotators (intra and inter-observer variability)
    - There may exist hard-to-annotate image parts and incorrect annotations as a result

17

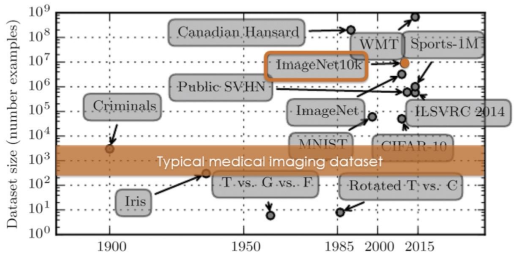
Common challenges

- Hard to define expressive features
  - Large variance may exist within samples
  - Noise and artifacts typically exist in samples/images due to non-ideal conditions in experimental setup and imaging
- Traditionally, features are manually defined based on domain-specific knowledge, human intuition, and known mathematical theories and tools
  - Sometimes, this process of extracting handcrafted features is not "*that effective*"
- Recently, deep learning has shown great promise as an alternative to employing handcrafted features
  - But it requires large datasets for training

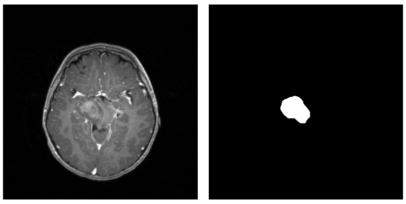
18

Common challenges

- Need of large annotated datasets to train deep models
  - **The more variety there is in the data, the larger the training dataset needs to be**
- Difficult to access to large high-quality annotated datasets
  - ImageNet is extremely powerful since it is huge and accurately annotated
- Imbalanced data problem



www.quantib.com/blog/deep-learning-radiology-and-challenges-radiology-ai

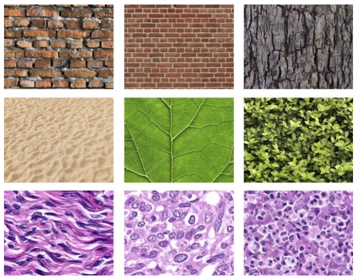


Separation of touching objects

What should I have learned by now?

1. General notion on medical imaging modalities
2. Preliminaries for an algorithm design
3. Common challenges in medical image analysis
4. **Traditional methods for image representation (feature extraction)**
5. Medical image classification
6. Medical image segmentation
7. General notion on other commonly studied problems
   - Such as image synthesis, image denoising, image registration, super-resolution imaging, and multiple instance learning

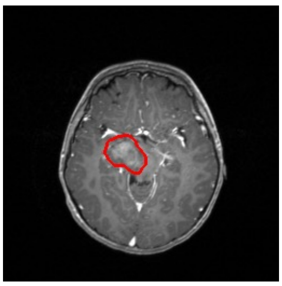
Handcrafted feature extraction for medical images

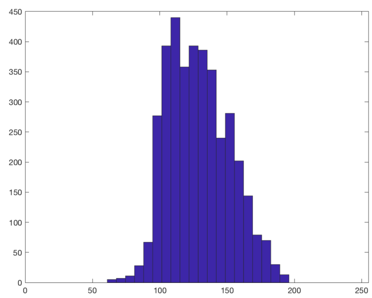


- Intensity features
  - Quantify color/grayscale distribution of pixels
  - First-order statistical texture
- Texture features
  - Similar structures repeated over and over again (repeated patterns)
  - Statistical approach is to quantify spatial arrangement of pixel intensities
  - Structural approach is to define texture on objects/primitives
- Structural features to quantify spatial distribution of objects/primitives
- Morphological features to quantify the shape/size of a segmented object

First-order statistical features

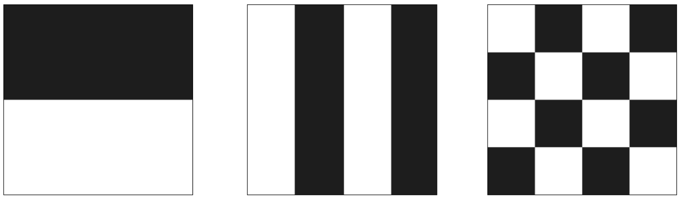
- Extracted based on histogram analysis
  - Mean
  - Standard deviation
  - Skewness
  - Kurtosis
  - Entropy
  - Min, max, quartiles
- More useful if you first identify regions of interest



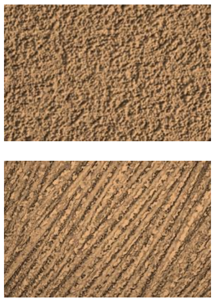


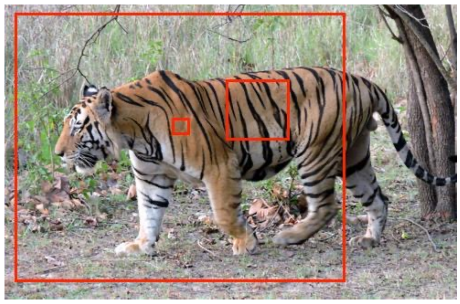
Texture

- Statistical approaches
  - Co-occurrence matrices
  - Run length matrices
  - Local binary patterns
  - Laws kernels
  - Gabor filters
  - …



Three images with the same intensity histogram, but different textures

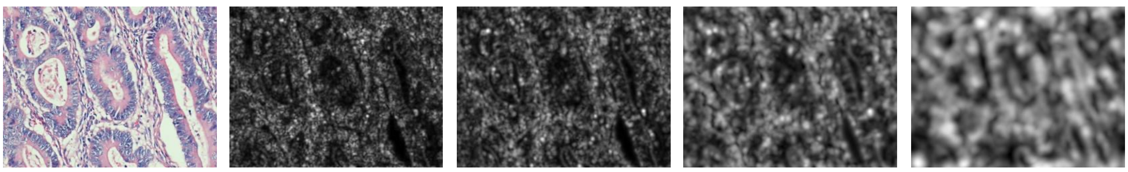




Texture depends on the scale at which it is viewed

Intensity/texture features to represent an entire image

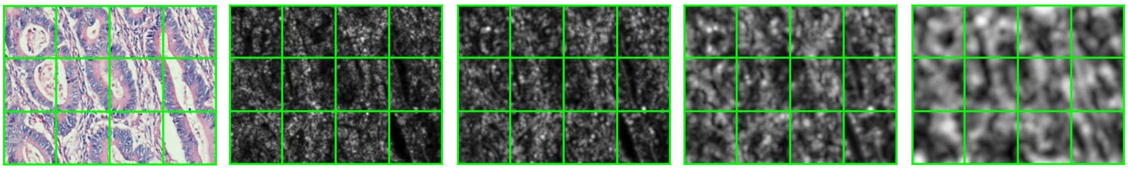
1. Calculate them on the entire image
   - e.g., Haralick features of the cooccurrence matrix calculated on all image pixels
   - e.g., first order statistical features (mean, standard deviation, entropy, etc.) of filter responses calculated on all image pixels



Example: The grayscale image is convolved with Gabor filters of six orientations and four scales. For each scale, the filter responses of different orientations are averaged to obtain rotation-invariant features. Statistical features will be separately calculated on each of these four average response maps.

Intensity/texture features to represent an entire image

2. Grid-based approach
   - Divide an image into grids
   - Calculate features on each grid
   - Aggregate the feature values

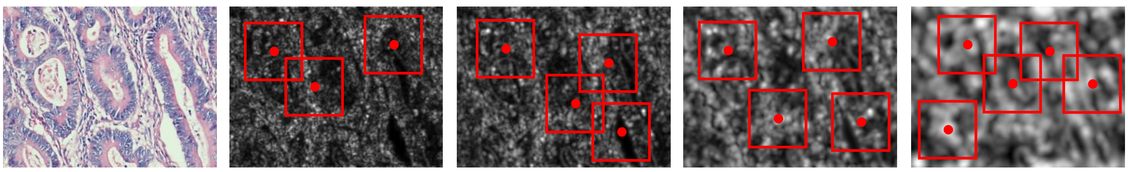


Example: The grayscale image is convolved with Gabor filters of six orientations and four scales. For each scale, the filter responses of different orientations are averaged. Then, for each grid entry of each average map, statistical features will be separately calculated. Afterwards, the statistical features will be aggregated for the same map.

25

Intensity/texture features to represent an entire image

3. Calculate them on the keypoints
   - Find the key points
   - Locate a window on each of these key points
   - Calculate features on each window
   - Aggregate the feature values

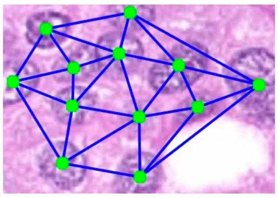


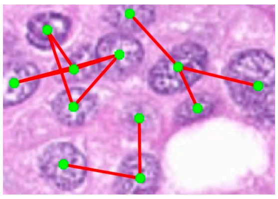
Example: The grayscale image is convolved with Gabor filters of six orientations and four scales. For each scale, the filter responses of different orientations are averaged. Then, for each window located on a keypoint, statistical features will be separately calculated. Afterwards, the statistical features will be aggregated for the same map.

26

Structural features

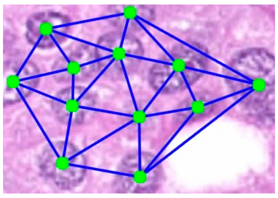
- Quantify spatial distribution of objects/primitives
  - By defining texture metrics on the objects/primitives
  - By making use of graphs
- To construct a graph on a medical image, you need to
  - First define its vertices
  - Then assign edges in between these vertices
- Features can be extracted at the local or the global level

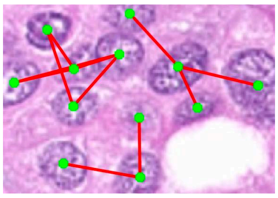




Structural features

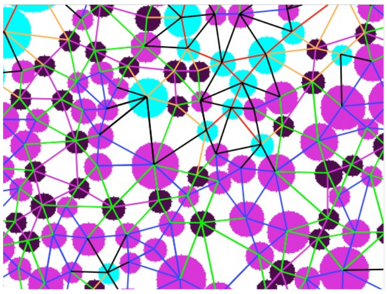
- *Local features* are to quantify connectivity information for each graph vertex
- *Global features* are to quantify connectivity information for an entire graph
  1. Statistics on the local graph features
  2. Using connected components
  3. Using graph spectrum (eigenvalues of the graph's adjacency matrix)
  4. Additional metrics such as
     - Statistics on the graph edge lengths
     - Statistics on the triangle areas, if Delaunay triangulation is used

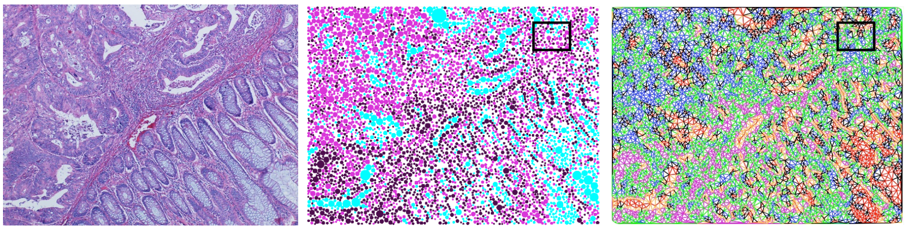




Example: Defining texture by graphs

- Represent an image with a graph of cytological tissue components
  - Approximate their locations with circles
  - Construct a Delaunay triangulation on the circles
  - Color each triangle edge according to the circle (component) type of its end-nodes
- Define texture descriptors on this color graph



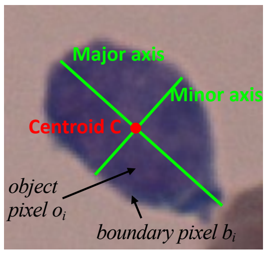


Tosun and Gunduz-Demir, "Graph run-length matrices for histopathological image segmentation," IEEE T Medical Imaging, 2011.

29

Morphological (geometric) features

- Mostly defined for an individual object (for each segmented region)
- Quantify the shape and size characteristics of this object (region)
- *Size*
  - *Area* is the number of pixels in the object
  - *Perimeter* is the number of pixels on the object's boundaries (based on 4-connectivity or 8-connectivity)
  - *Major axis* is the longest line that goes through the centroid
  - *Minor axis* is the line that is perpendicular to the major axis and goes through the centroid
  - *Average radius* is average length of the radial lines from the centroid to every boundary pixel



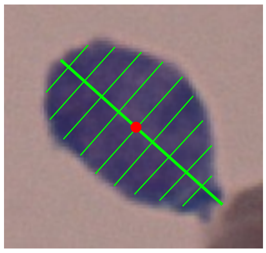
$$C = \frac{1}{N}\sum_{i=1}^{N} coord(o_i)$$

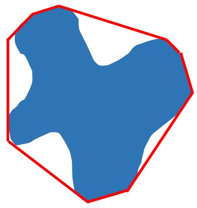
$$Avg\ radius = \frac{1}{N}\sum_{i=1}^{N} \|b_i - C\|$$

30

Morphological (geometric) features

- *Shape*
  - *Circularity* $= 4\pi \frac{area}{perimeter^2}$ (0 < metric <=1, 1 for circles)
  - *Symmetry* is based on the length difference of two segments on the same line perpendicular to the major axis
  - *Smoothness* is the sum of the smoothness of the boundary pixels
    - For each boundary pixel, it is the difference between its radius (radial line from the centroid to this pixel) and the average of the radii of its surrounding (or close enough) boundary pixels
  - *Eccentricity* is the ratio of the minor to the major axis length





- Can be quantified also using the convex hull of the object

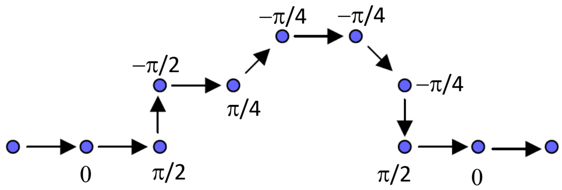
$$Convexity = \frac{perimeter\ (object)}{perimeter(convex\ hull)} \qquad Solidity = \frac{area\ (object)}{area(convex\ hull)}$$

(0 < metric <=1)

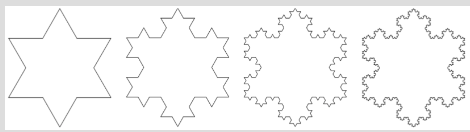
31

Morphological (geometric) features

- *Boundary descriptors*
  - *Curvature* is to quantify the amount by which the boundary contour deviates from being a straight line
  - *Fractal dimension* of the boundary contour quantifies the rate at which its length (perimeter) increases as the measurement scale decreases



Fractals exhibit similar patterns at different scales
This property is known as *self-similarity*



Koch snowflake and cauliflower as examples of fractals

32

What should I have learned by now?

1. General notion on medical imaging modalities
2. Preliminaries for an algorithm design
3. Common challenges in medical image analysis
4. Traditional methods for image representation (feature extraction)
5. **Medical image classification**
6. Medical image segmentation
7. General notion on other commonly studied problems
   - Such as image synthesis, image denoising, image registration, super-resolution imaging, and multiple instance learning

33

Medical image classification

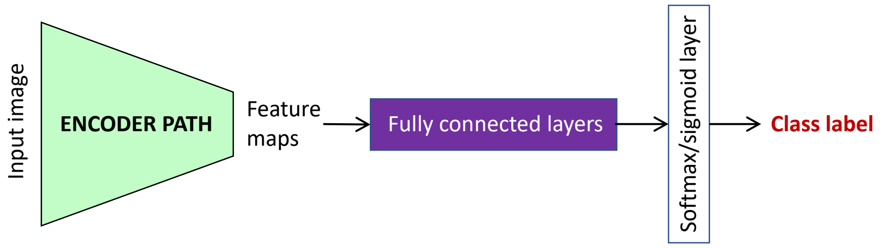
- Traditional classification approaches
  - Images are classified based on their handcrafted features using linear discriminants, neural networks, support vector machines, decision trees, …
  - There may exist some issues in practice
    - Unbalanced class distributions and small training sets (consider augmenting data)
    - Features with different orders of magnitudes or high values (consider normalization)
    - Overfitting or underfitting (consider regularization techniques, use "*better*" designs, use "*proper*" validation techniques)

- Deep learning models
  - Convolutional neural networks (CNNs)
  - Transfer learning

34

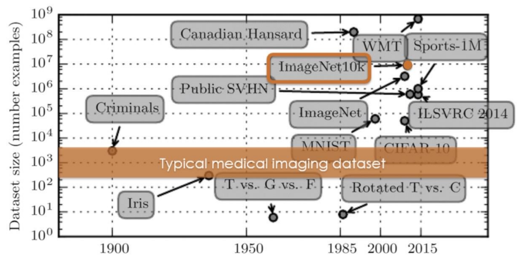
CNNs for image classification

- A CNN compresses an image into a set of feature maps to capture semantic/contextual information from the image
- This compression corresponds to downsampling the image using convolution and pooling layers
- Then it puts fully connected layers on the top of the feature maps to predict a class for the entire image



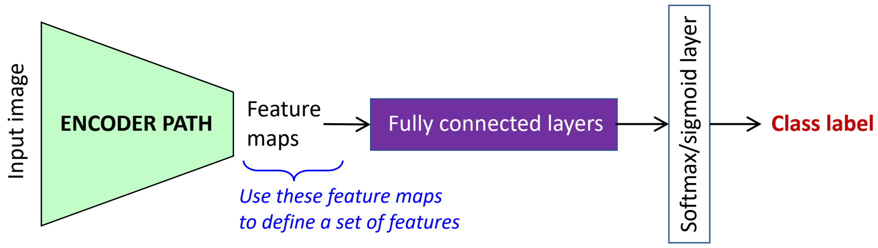
Transfer learning

- Dataset sizes are typically small
- Thus, it is popular to use transfer learning, which employs networks (and thus, their learned weights) previously trained on large datasets
- Two main approaches
  1. Use a pretrained network as a feature extractor
  2. Finetune a pretrained network on the medical data

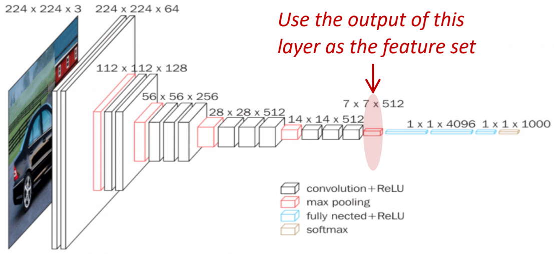


www.quantib.com/blog/deep-learning-radiology-and-challenges-radiology-ai

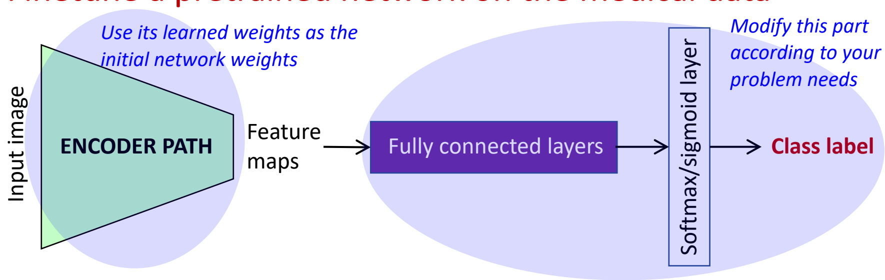
Use a pretrained network as a feature extractor



1. Feed your image to the pretrained network (may need to resize the image)
2. Calculate feature maps using the learned weights
3. Use the feature maps at the last layer of the encoder to define a feature set for your image (also possible to use feature maps of the other layers)
4. Use a classifier on this feature set

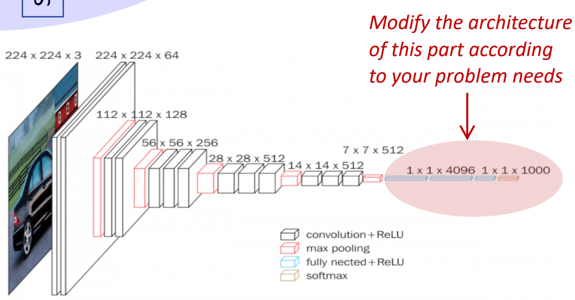


Finetune a pretrained network on the medical data

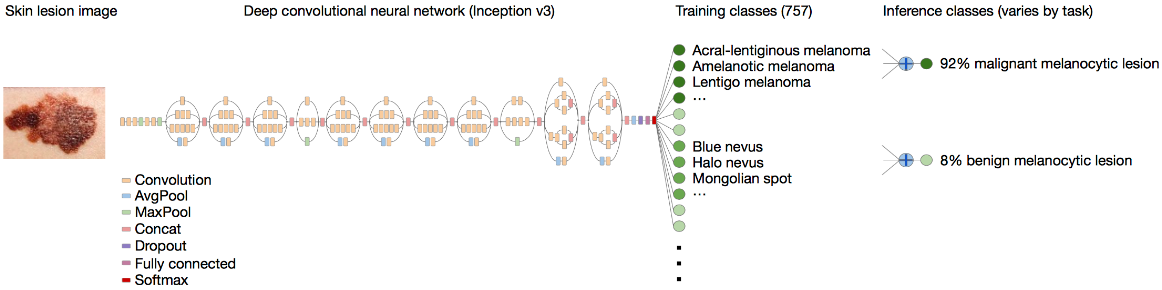


If you want to design your own network architecture, you need to train it from scratch

1. Feed your image to the pretrained network (may need to resize the image)
2. Modify the network architecture after the encoder path. Changing the last softmax/sigmoid layer according to your classification problem is a must. May also need to change the fully connected layers.
3. Use the learned weights (of the encoder) as the initial networks weights
4. Finetune the weights by backpropagation on your own medical data



Example: CNN for skin cancer classification



The authors used the Google Inception v3 CNN architecture pretrained on the ImageNet dataset (1.28 million images over 1,000 generic object classes) and finetuned on their own dataset of 129,450 skin lesions. They resized each image to 299x299 pixels to make it compatible with the original dimensions of the Inception v3 network architecture. They defined 757 training classes, for which the probabilities would be accumulated to infer the final inference class.

Esteva et al, 2017. Dermatologist-level classification of skin cancer with deep neural networks. Nature 542, 115–118. https://www.nature.com/articles/nature21056

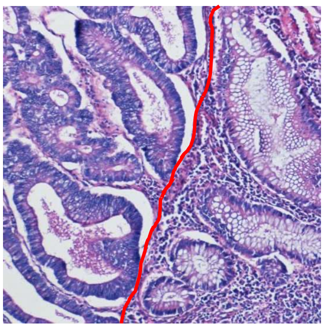
39

What should I have learned by now?

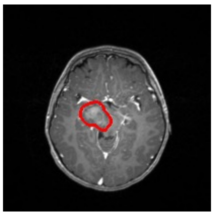
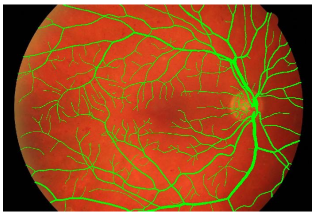
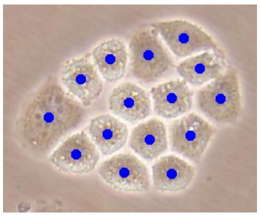
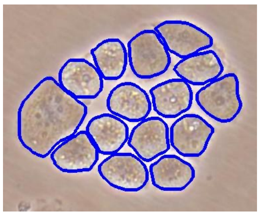
1. General notion on medical imaging modalities
2. Preliminaries for an algorithm design
3. Common challenges in medical image analysis
4. Traditional methods for image representation (feature extraction)
5. Medical image classification
6. **Medical image segmentation**
7. General notion on other commonly studied problems
   - Such as image synthesis, image denoising, image registration, super-resolution imaging, and multiple instance learning

40

Medical image segmentation



- Typically deals with
  - Locating biological structures (cells, glands, vessels, tumor, lesions, etc.) on an image
    - ➢ Similar to object detection in other domains
    - ➢ Could be in the form of detecting approximate object locations (e.g., centroids, bounding boxes) or finding their exact boundaries
  - Dividing a heterogeneous image into its homogenous regions
    - ➢ Similar to scene segmentation in other domains

41

Medical image segmentation

- Traditional segmentation algorithms
  - Histogram-based, clustering-based, region growing (watersheds), graph-based, …
  - Preprocessing is typically required (filters)
  - Postprocessing and refinements may also be necessary (mathematical morphology, active contour models, …)
- Deep learning models
  - Convolutional neural networks
  - Dense prediction networks

1. Classifying each pixel in a sliding window fashion, used by earlier studies, is expensive as it requires lots of redundant calculations

2. Trade-off between localization accuracy and the use of context
- *Larger patches require more max-pooling layers that reduce the localization accuracy*
- *Small patches results in seeing only little context*

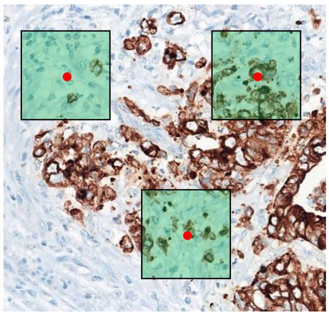
Dense prediction networks, used by recent studies, have greatly improved efficiency and accuracy.

42

CNNs for object detection and segmentation

Training:

- Small patches are cropped around individual pixels
- Each patch is labeled with the class of the pixel, around which it is cropped
- CNN is trained on these small patches



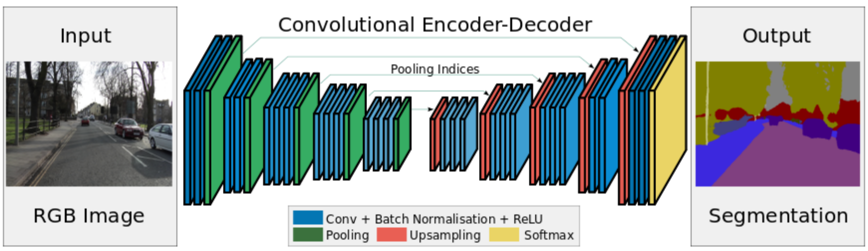
Detection/segmentation:

- For an entire (large) image, patches are obtained using a sliding window approach
- These patches are classified by the trained CNN
- Outputs (i.e., posteriors) generated by this CNN are commonly postprocessed

43

Dense prediction networks

- They recover a larger-size segmentation map from the compressed image
  - Downsampling path captures semantic/contextual information
  - Upsampling path recovers spatial information
  - No fully connected layer is used on the top
  - Skip connections (concatenations) from downsampling to upsampling layers are often used to recover the fine-grained spatial information lost in the downsampling path



44

U-Net architecture

- Long skip connections are defined between an encoder layer and the corresponding decoder layer
  - Deconvolution is applied on the concatenation of high resolution features of an encoder layer and the output of the previous upsampling layer
  - Helps better recover the fine-grained spatial information lost in downsampling
  - U-shaped architecture with an equal number of downsampling and upsampling layers

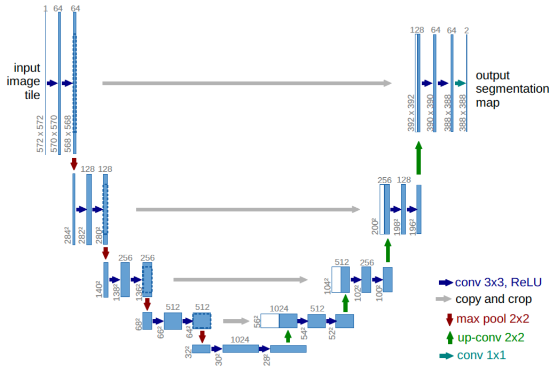


Fig. 1. U-net architecture (example for 32x32 pixels in the lowest resolution). Each blue box corresponds to a multi-channel feature map. The number of channels is denoted on top of the box. The x-y-size is provided at the lower left edge of the box. White boxes represent copied feature maps. The arrows denote the different operations.

Ronneberger et al, 2015. U-net: Convolutional networks for biomedical image segmentation. MICCAI 2015.
https://arxiv.org/pdf/1505.04597.pdf

45

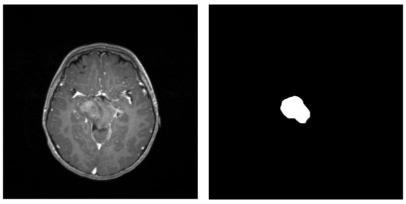
Loss functions

- Standard definition

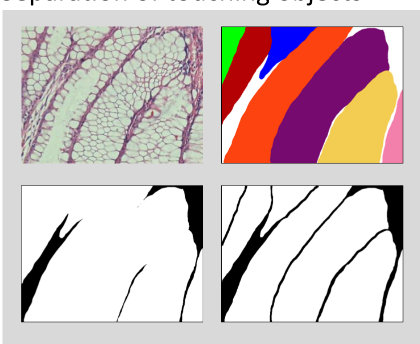
$$loss = \sum_{I \in D_{tr}} \sum_{p \in I} loss(y_p, \hat{y}_p)$$

- Imbalance distribution of positive and negative pixels as well as the difficulty of separating touching objects may require using custom definitions
  - Weighted loss, focal loss, Dice loss, Tversky loss, shape-aware loss, adaptive loss adjustments, …

Imbalanced classes



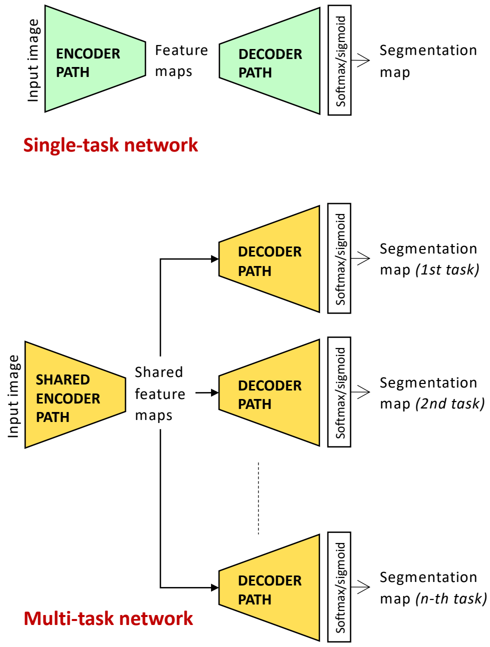
Separation of touching objects



46

Multi-task networks

- Dense prediction networks that learn related but different tasks from shared feature representations
- They consist of a shared encoder path and multiple decoder paths, one defined for each task
- Joint loss is defined usually as a linear combination of losses defined on all tasks
- All tasks are concurrently learned in parallel by training the network to minimize this joint loss
- This approach helps better avoid local optimal solutions as it is less likely to finetune the weights of the shared encoder for all tasks at the same time



Contour-aware network

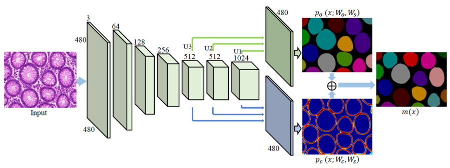
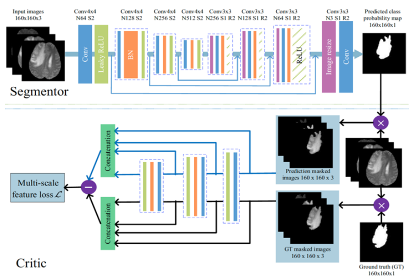


Figure 3: The overview of the proposed deep contour-aware network.

Chen et al., 2017. DCAN: Deep contour-aware networks for accurate gland segmentation.
https://www.sciencedirect.com/science/article/pii/S1361841516302043

Segmentation networks with adversarial loss



Xue et al., 2018. SegAN: Adversarial network with multi-scale L1 loss for medical image segmentation.
https://link.springer.com/content/pdf/10.1007/s12021-018-9377-x.pdf

Image reconstruction as an auxiliary task

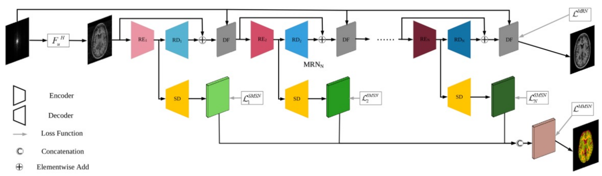


Figure 3: The SegNetMRI structure, formed by connecting the discussed MRN (top) for reconstruction, with MSN (bottom) for segmentation.

Sun et al., 2018. Joint CS-MRI reconstruction and segmentation with a unified deep network.
https://arxiv.org/pdf/1805.02165.pdf

What should I have learned by now?

1. General notion on medical imaging modalities
2. Preliminaries for an algorithm design
3. Common challenges in medical image analysis
4. Traditional methods for image representation (feature extraction)
5. Medical image classification
6. Medical image segmentation
7. **General notion on other commonly studied problems**
   - **Such as image synthesis, image denoising, image registration, super-resolution imaging, and multiple instance learning**

49

Example: Synthetic data augmentation

- Unconditional image synthesis by a GAN to augment training data for liver lesion classification in CT images
- The generator network inputs a vector of 100 random numbers drawn from a uniform distribution and outputs a liver lesion image of size 64x64

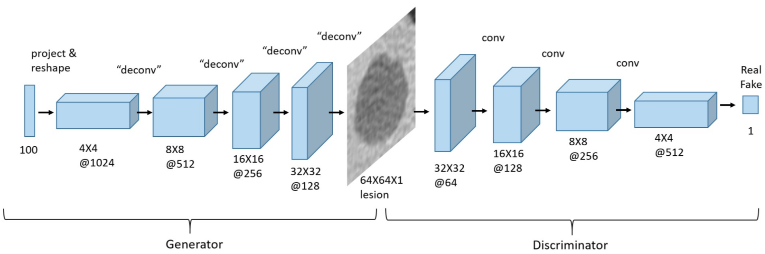


Fig. 2: Deep Convolutional GAN Architecture (generator+descriminator).

Frid-Adar et al., 2018. Synthetic data augmentation using GAN for improved liver lesion classification.
https://arxiv.org/abs/1801.02385

50

Example: CT synthesis from MR

- Computed tomography requires a patient to expose radiation whereas MRI does not
- Learns to synthesize CT images by training a conditional GAN on pairwise aligned MR and CT images

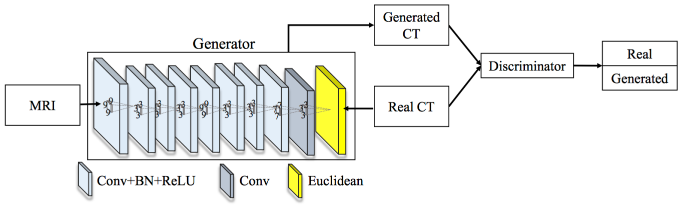


Fig. 2. Architecture used in the Generative Adversarial setting used for estimation of synthetic images.

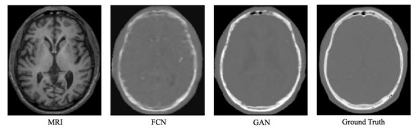


Fig. 4. Visual comparison for impact of adversarial training. FCN means without adversarial training, and GAN means with adversarial training.

Nie et al., 2017. Medical image synthesis with context-aware generative adversarial networks.
https://arxiv.org/pdf/1612.05362.pdf

51

Example: Low-dose CT denoising

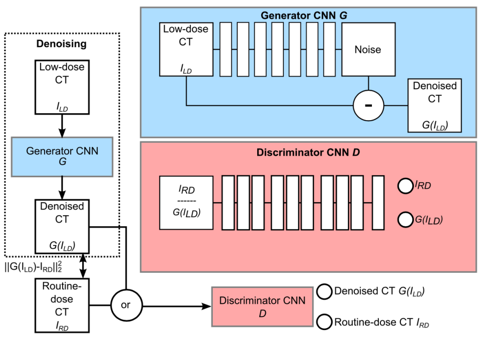
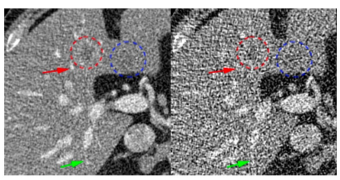


Fig. 1. Overview of the proposed pipeline for noise reduction in low-dose CT. The generative adversarial network consists of two components: a generator CNN and a discriminator CNN. The generator uses regression to determine the routine-dose HU value at every voxel in a low-dose CT. It does this through a skip connection which subtracts an estimated noise image from the input low-dose image. The discriminator tries to distinguish reduced noise CT images from real routine-dose images.

- To learn the appearance of routine-dose CT images from low-dose CT images
- It does not require spatially aligned pairs of low-dose and routine-dose images



(a) Full Dose FBP (b) Quarter Dose FBP

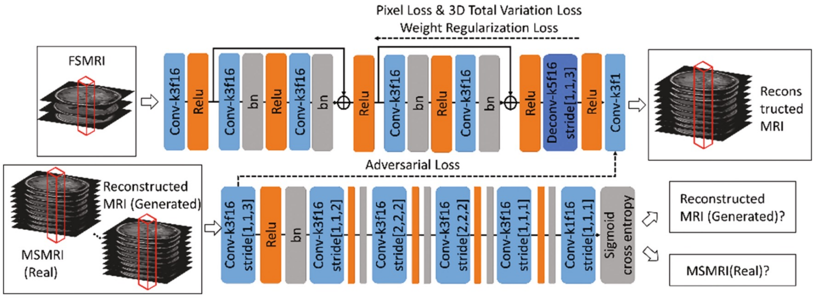
Yang et al., 2018. Low-dose CT image denoising using a generative adversarial network with Wasserstein distance and perceptual loss.
https://arxiv.org/pdf/1708.00961.pdf

Wolterink et al., 2017. Generative adversarial networks for noise reduction in low-dose CT.
https://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=7934380&tag=1

52

Example: Super-resolution imaging

- To reconstruct MR images with thinner slice thickness from regular thick slice images
  - Thinner slices have higher spatial resolution and provide more diagnostic information, but also have higher imaging cost both in time and expense
- Generator with residual connections
  - Minimizes MSE loss function, L2 regularization loss, 3D total variation loss (the sum of MSE between every neighbor slices of the reconstructed images), and adversarial loss



Li et al., 2017. Reconstruction of thin-slice medical images using generative adversarial network.
https://link.springer.com/content/pdf/10.1007%2F978-3-319-67389-9_38.pdf

Example: Registration

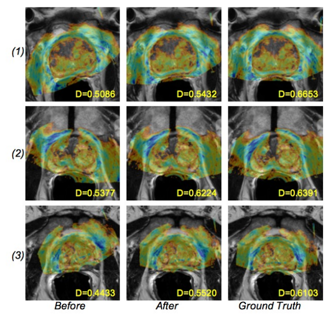
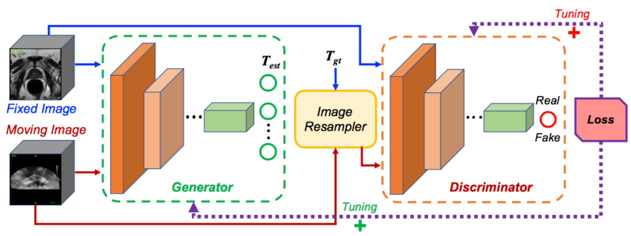


Fig. 2: Example registration results from 3 different cases. MR images are shown in gray level and corresponding TRUS images are superimposed in pseudo color.

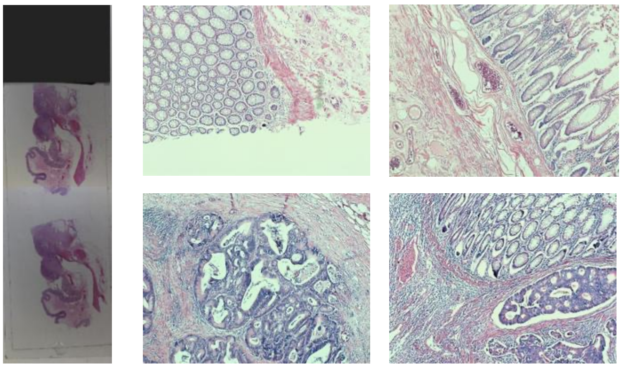
- To align MR and transrectal ultrasound (TRUS) images
- Generator network directly estimates transformation parameters between the input MR and TRUS image pairs
- Image resampler uses either the estimated or the ground truth transformation to interpolate a moving image
  - Parameters: rotation in [-25,25] degrees and translation in [-5,5] mm
- Discriminator network tells whether an input pair is aligned using the estimated or the ground truth transformation

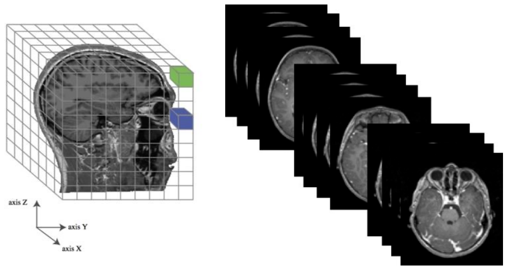


Yan et al., 2018. Adversarial image registration with application for MR and TRUS image fusion.
https://arxiv.org/pdf/1804.11024.pdf

Multiple-instance learning

- Although obtaining ground-truth local annotations (for patches or pixels) is costly, time-consuming or sometimes even impossible, ***global labels are available for whole images***





These global labels do not apply to all patches or pixels inside a whole image

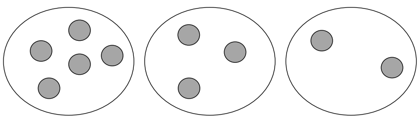
55

Multiple-instance learning

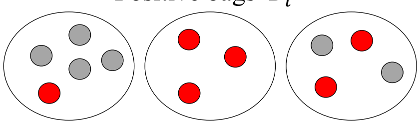
- Multiple-instance learning (MIL) is an extension of supervised learning to train learners using such weakly labeled data (i.e., only global labels are available)
- Training instances are arranged in sets (called *bags*) and a label is provided for an entire bag (not for every instance separately)
- In binary setting, a bag is labeled as
  - Positive if there is at least one positive instance in the bag
  - Negative if all instances in the bag are negative

The main challenge is to cope with not knowing which instances in a positive bag are actually positive and which are negative

Negative bags B_i^-



Positive bags B_i^+



56

Example: MIL in digital mammography

- A bag is defined for each mammogram which is adaptively partitioned into regions (each corresponds to an instance)
- Each bag is labeled as positive if the corresponding mammogram contains an abnormal lesion (no region-based labeling)
- Each region (instance) is quantified with a set of features
  - Extracted using mass and microcalcification detection algorithms
  - Well known texture features (those defined on gray-level co-occurrence matrices, gray-level run-length matrices, and local binary patterns)
- Various MIL algorithms are used to classify a mammogram as well as to identify the most abnormal region(s) in the mammogram
  - Diverse density (DD), axis-parallel rectangles (APR), multiple-instance support vector machines (mi-SVM and MI-SVM), and MILBoost

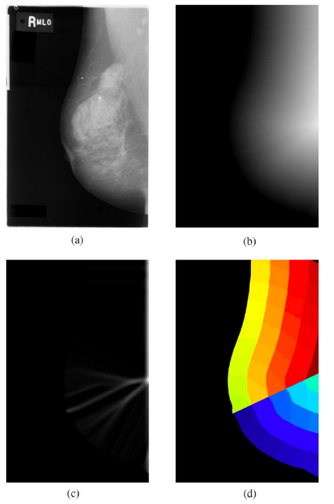


Fig. 1. Breast segmentation and region definition (a) original optical density image (b) distance transform (c) distance transform ridges (d) region segmentation.

Quellec et al., 2016. Multiple-instance learning for anomaly detection in digital mammography.
https://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=7390250

57

Example: MIL and deep learning in pathology slides

- A bag is defined for a histopathological image and instances are pixels in the image
- Labels are available at the image-level (no pixel labels are available)
- A baseline FCN is trained to predict pixel labels
- **DWS-MIL (multiple instance learning framework with deep weak supervision)** aims to control and guide the learning process across multiple scales
- Overall loss defined on all side-output layers associated with different scales

$$\mathcal{L}_{mil} = -\sum \left(\boldsymbol{I}(Y_i = 1) \log \widehat{Y}_i + \boldsymbol{I}(Y_i = 0) \log(1 - \widehat{Y}_i) \right)$$

$$\widehat{Y}_i = \left(\frac{1}{|X_i|} \sum_{k=1}^{|X_i|} \widehat{Y}_{ik}^r \right)^{1/r}$$

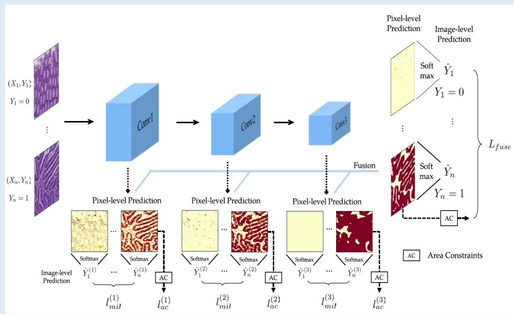


Fig. 3. Overview of our framework. Under the MIL setting, we adopt first three stages of the VGGNet and connect side-output layers with deep weak supervision under MIL. We also propose area constraints to regularize the size of predicted positive instances. To utilize the multi-scale predictions of individual layers, we merge side outputs via a weighted fusion layer. The overall model of equation (13) is trained via back-propagation using the stochastic gradient descent algorithm.

Jia et al., 2017. Constrained deep weak supervision for histopathology image segmentation
https://ieeexplore.ieee.org/stamp/stamp.jsp?arnumber=7971941

58

Thank you!

59