Group_12 (MSDS20005 (Muhammad Ali), MSCS20036(Aqsa Arshad))

Q1: Why does Graph Neural Network Gives better image segmentation for medical images than other pixel or traditional ML approaches?

Let us first discuss pixel level ML approaches starting from point distribution model upto Cnn, All of these works on pixel space directly. For these to work, we need to have accurate pixel labels. Most of these work on minimum pixel level loss for segmentation.

What they lack is landmark annotations and connectivity structure. A Lot of segmentation tasks in medical images consist of structured data and can be represented as landmarks. In traditional ML approaches for medical images we aren't taking the connectivity of structures.

GNN are structural models that appear as a natural way to represent landmarks, contours and surfaces. That's why they perform a lot better than other ML approaches as they also take in the connectivity of the objects while training.

Q2: What is image occlusion and why is it a big issue in medical imaging?

Occlusion means the blockage. Image occlusion means anything that hides some part or complete image. It is quite a big issue in medical imaging as the process of capturing the internal organs of humans involves radiation or waves. The cheapest and commonly used method is using x-rays. That captures a 3d image of the objects into a 2d paper, called film (we are referring to the x-ray process.). Many organs that overlap each other generate occlusion for the object of interest.

This printed thin film that can be easily messed up. In Fact sometimes small movement during capturing an image can generate occlusion. It greatly affect our segmentation solution in medical images

Q3: What is atlas-based segmentation? Explain how this idea (Atlas Based Segmentation) can be extended to multi-atlas and why is it important in medical field

(A) Let's first define an atlas. An atlas is a reference image that has been accurately segmented, that you can consider to be a reference or template for all other segmentations. Standard procedure for atlas-based segmentation is as follow

- Identify a reference image and manually segment the image, as accurately as possible. This is your atlas.
- For a new image that needs to be segmented, first register (align) it to the atlas

● Copy the atlas labels onto the image that needs segmenting

(B) In multiple atlas segmentation of images, we have multiple atlas, let's say 100. We match the image through the most similar one and pick the top 5 of them. Then we map them to the original image to these atlas and take the average of these atlas.

This is quite useful in medical images as different doctors would do different segmentation of the same image

Q4 What is spectral space and spatial space? What is spectral convolution and why do we use spectral Convolution in graphs rather than simple convolution while dealing with medical images?

(A) Let us first discuss Spatial space which means Treating images as they are (as pixels). The value of the pixels of the image changes with respect to the scene.

If we talk about spectral space, then we should know about the image spectrum first(which means that Image Has been converted to spectral space).

Spectrum means frequencies or related quantities such as energies, eigenvalues, etc. it is also called frequency domain. In the frequency domain, we deal with the **rate at which the pixel values are changing** in the spatial domain. This is also called Fourier space.

While "spectral" may sound complicated, for our purpose it's enough to understand that it simply means decomposing a signal/audio/image/graph into a combination (usually, a sum) of simple elements (wavelets, graphlets).

(B) Convolution in spatial space is just Multiplication in Frequency Domain. So we can say that spectral convolution means multiplying with the Laplacian matrix (which is equal to L=D-A; where D is diagonal matrix and A is adjacency matrix).

(C) Spectral convolution is important in graphs as this involves the Eigen decomposition that helps us in understanding the underlying structure of the graph with which we can identify clusters/sub-groups of this graph.Unlike Spectral Convolution which takes a lot of time to compute, Spatial Convolutions are simple and have produced state of the art results on graph classification tasks.

Group 7: Hassan Khalid MSDS20002

Q1:Why is Graphical Super Resolution required in the medical field?
Answer: Some of the medical information extracted from different modalities like MRI is in graphical format, and therefore, the super-resolution of the same required graphical super-resolution algorithms.
The most common example in brain connectomes (is a comprehensive map of neural connections in the brain).

Q2: **What is the significance of the use of loss for addressing the issue of global HR topology in reference to the medical domain?**

Answer: The high-resolution information acquired from the high-quality medical image modalities has a certain type of typological characteristics associated with it, and so the super-resolved result must also preserve the same for the better diagnosis at the later stage and for the improvement of the overall accuracy of the algorithm; generating realistic data.

Q3: Why is Super Resolution required in the medical field?

Answer: The high-resolution medical modalities provide fine-grained information, and the same can be used for a better diagnosis at the later stages. However, high-quality data generating modalities are scarce and costly, therefore super-resolution algorithms are required in the medical domain.

Group 3: (Hamza Munir MSDS20061 and Mehmood Ali MSDS20090)

Why do we need domain adaptation in medical imaging?

As we know in the field of medical imaging publicly available dataset is a big issue. We don't have a large amount of dataset available. Some of the reasons for this are patients' privacy, data capturing cost, high machinery cost, data maintenance cost, etc. The labeling of the data is also very expensive and technical, as we cannot get label data from some random person, so we need experts of the domain to label the data. On the other hand, state-of-the-art deep learning models require large amounts of data, for example Deep Neural Network Models. So, domain adaptation is very important in medical imaging, we can learn from data from other domains and try to optimize the problem of another domain.

For Example, in the case of Surgical Videos, we can train the model on VR (Virtual Reality) videos and use the model in a clinical environment.

What is Hard domain Adaptation and Soft Domain Adaptation? Which is more useful in Medical Imaging?

For example, we are trying to build a network to identify some tasks in clinical videos using Domain Adaptation Technique with a small number of Clinical Videos and a large number of Virtual Reality Videos. In case of Soft Domain Adaptation, we will use both (Clinical and VR videos) datasets and try to accurately classify tasks in clinical videos. In Hard Domain Adaptation, we will use only Virtual Reality Videos and make models able to accurately classify clinical videos.

Hard domain adaptation is difficult to use in medical imaging as we need near 100 percent accuracy in medical related problems and we cannot achieve it without adding targeted domain data in training of the model. So, the soft domain adaptation is helpful in this case and achieves good results as soft domain adaptation caters data from both domains.

Transfer learning is helpful or not in medical imaging?

Transfer Learning is helpful in some cases and not helpful sometimes in medical imaging. Transfer learning models are trained on the large pools of data with thousands of different classes.

In case, if we have a large amount of data and similar images or features are already learned by the pre-trained model, we can easily use the pre-trained model. In case, if we have small data and pretrained models haven't learned the similar data, so we cannot use pretrained models. In general cases we can freeze and use higher layers of pretrained models and retrain the ending convolutional and dense layers to fine tune the model at scale.

Group 5: Momina Malik and Maria Iqbal

How can we discard the image with the lower quality, specifically retinal images?

We can check the quality of images using four common quality indicators, including blurring, uneven illumi-nation, low-contrast, and artifacts.If the image doesn't have any of these it is a good image. But if the image has some slight low-quality indicators,which can not observe the whole image clearly (e.g., low-contract and blur) or affect the automated medical analysis methods (e.g., artifacts), but the main structures (e.g., disc, macula regions) and lesion are clear enough to be identified by ophthalmologists then it is a usable one. Otherwise we can say the image quality is very bad and we can reject it.

What are existing methods which are currently working in the market?

Recently, several methods for RIQA specifically have been proposed, which can be divided into two main categories: structure-based methods and feature-based methods. Structure-based methods employ segmented structures to determine the quality of retinal images. For example, an image structure cluster-ing method was proposed to extract compact representations of retinal structures to determine image quality levels. Blood vessel structures are also widely used for identifying the quality of retinal images. However, structure- based methods rely heavily on the performance of structure segmentation, and cannot obtain latent visual features from images. Feature-based methods, on the other hand, directly extract feature representations from images, without structure segmentation. For example, features quantifying image color, focus, contrast and illumination can be calculated to represent the quality grade. Wang et.al. employed features based on the human visual system, with a support vector machine (SVM) or a decision tree to identify high-quality images.

Why Retinal image quality assessment (RIQA) is essential?

Retinal image quality assessment (RIQA) is essential for con-trolling the quality of retinal imaging and guaranteeing the reliability of diagnoses by ophthalmologists or automated analysis systems.Retinal images are widely used for early screening and diagnosis of several eye diseases, including diabetic retinopathy (DR), glaucoma, and age-related macular degeneration (AMD). However, retinal images captured using different cameras, by people with various levels of experience, have a large variation in quality. Existing RIQA methods focus on the RGB color-space and are developed based on small datasets with binary quality labels (i.e., 'Accept' and 'Reject').we first re-annotate an Eye-Quality (EyeQ) dataset with 28,792 retinal images from the EyePACS dataset, based on a three-level quality grading system (i.e., 'Good', 'Usable' and 'Reject') for evaluating RIQA methods.

Group 13: MSCS21005 and MSDS20049

Learning Conditional Deformable Shape Templates for Brain Anatomy

Q1)   Why is a single and fixed brain template not sufficient in neuroimaging (neuroimaging is a discipline that studies the structure and function of the nervous system by means of imaging technology) analysis?. What factors can affect the brain structure?

A brain template is a powerful tool that enables the analysis of geometric variability of a brain. A brain template can be chosen as one of the brain MRI images in a given dataset. A single and fixed template can struggle to accommodate complex structural differences across a heterogeneous group of individuals, and can lead to biased and misleading analyses thats' why a deformable brain template(A deformable template is an image that can be geometrically deformed to match images in a dataset, providing a common reference frame) is required. Demographics(different regions of the country),

clinical(relating to the observation and treatment of actual patients rather than theoretical or laboratory studies), or other confounding factors(A confounding factor is an outside influence that changes the effect of a dependent and independent variable) can influence the shapes and sizes of brain regions.

Q2) What do you mean by a conditional template? How is this template helpful in studying the disease?

Conditional templates capture important trends related to attributes i.e. age, gender, Alzheimer's disease (Alzheimer's is a type of dementia that affects memory, thinking and behavior), and are useful for dealing with confoundersA confounders are an outside influences that changes the effect of a dependent and independent variable). For example, in studying disease impact, for some tasks it may be helpful to register scans to age-specific templates rather than one covering a wide age range.

Q3) What is the purpose of Transpose Convolution in brain anatomy?

Brain anatomy is the identification and description of the structure of the brain. Some structures of the brain are blurry and imprecise(low-resolution). In order to convert the low resolution image of the brain into high resolution(less blurry), transpose convolution is used. Transposed Convolutions are used to upsample the input feature map to a desired output feature map using some learnable parameters.

Group 6: Aiman Younas msds20087@itu.edu.pk Qazi Danish Ayub

Supervised Multi-topology Network Cross-diffusion for Population-driven Brain Network Atlas Estimation

Q1: How to visualize brain anatomy using graphs?

Graph theory provides a theoretic framework to represent the whole brain as a brain network and analyzing its topology. This network has associations and connections between various regions and subregions in the brain and can be defined as a graph G(V, E) , where nodes (V) may represent voxels or ROIs.For example, the Automated Anatomical Labeling (AAL) atlas divides the whole brain into 116 distinct brain structures (90 cortical and subcortical regions and 26 cerebellar regions).For each region, the representative time course in each region was obtained by averaging the time series across all voxels in the region. Then, the weight of each

edge (e ∈ E) in the brain network was computed based on the correlation coefficient between time series from all pairs of regions.

Q2: How to use topological measurements to determine important regions of the brain?

Recent research has shown that graph theory based complex network analysis provides a powerful framework for examining the topological properties of brain networks, where nodes represent brain regions, and edges represent the functional connections between brain regions .Topological properties of brain networks can be derived at the large-scale of the whole brain, i.e. metrics on the entire graph; at the intermediate scale of several regions of the brain. It follows that the selection of the topological metrics is an important step that requires great attention. Depending on the nature of the neuroimaging experiment, the functional connections, and the thresholding method, some graph indices can result in being more appropriate than other ones.

Q3: How to distinguish between a healthy and a non-healthy brain?

Graph matching is an essential step for inference and classification analyses on brain graphs. Graph matching is the process of finding a correspondence between the nodes and the edges of two or more graphs that satisfies some (more or less stringent) constraints ensuring that similar substructures in one graph are mapped to similar substructures in the others. In realistic applications, the matching is done by comparing the model graph which is stored in a file and an input graph that is given at the run time.

Group 20:

Muhammad Abdullah Nezami bscs19026

Hafiz Umair Ahmad bscs19064

Umais Muhammad bscs19035

Question 1: Why do we need skip-connections in u-net architecture?

Answer: While using U-Net we convolve an image multiple times and during this convolution, we lose some important low-resolution features of the image. So, we use skip connections which skip some layers of the network and transfer the features to the decoder. Skip-connections in U-Net are used to save the important low-resolution features of the image.

Question 2: How can we Train the model using patch extraction and why is it needed?

Answer: In Medical Imaging there are many models that are needed to be trained on a large dataset to have a good precision and less error. But collecting a large dataset is not easy so

patch extraction is a method used to make a small dataset large enough to train the model. In the patch extraction method an image is divided into a certain number of patches and different methods are used to extract the patches. Sometimes we extract them seamlessly and sometimes we use a stride of size N to extract them. Data is further Augmented by rotating the patches to 90°, 180° or 270°. Then the data is shuffled and fed to the model for training.

Question 3: Why do we use Contrast Limited Adaptive Histogram Equalization instead of Adaptive Histogram Equalization?

Answer: Adaptive histogram equalization is a contrast enhancement method for both natural and medical images, which computes several histograms corresponding to distinct sections of the image and uses them to redistribute the lightness values of the image. However, Adaptive histogram equalization tends to over-amplify the contrast in near-constant regions of the image and over- enhances noise. Contrast Limited Adaptive histogram equalization reduces the noise amplification problem by applying limitations on contrast amplification.

Group 22: BSCS19070, BSCS19003, BSCS19027

   1. **Why is Topological Coherence Important in Blood Vessel Segmentation?**
Ans: It is important in blood vessel segmentation because topological mistakes in structures like blood vessels, such as missing segments or inadequate merging in the segmentation might make it difficult to extract and characterize vessel pathways. This puts at risk the application of vessel segmentation that uses correct vessel pathways or vessel graphs. Topological mistakes might lead to incorrect results in these applications.

   2. **Previous models for blood vessel segmentation used pixel wise binary cross entropy to find loss, this increased pixel wise accuracy but topological mistakes were largely ignored leading to the model being prone to topological errors. How can we make the model learn to predict correct topology?**
Ans: The problem fixed in this model was that topology was ignored in previous models and pixel wise accuracy was targeted. This problem can be solved by penalizing the model on topological mistakes. We can cascade another network at the end of the segmentation network and this second network is responsible for learning the topology of the blood vessels and fixing all the topological mistakes in the segmentation output. The second network is penalized for topological mistakes and improves the topology of the final segmentation result.

   3. **How Does Cascading a Segmentation model and a Denoising variational auto encoder helps in obtaining blood vessel segmentation that are more topologically coherent and perform better than a single Segmentation Network?**

Ans: The Segmentation model first detects the high and low level features from the blood vessel images and produces a result. This result is considered as an input with hidden noise to the variational auto encoder that acts as a denoising model by learning noise patterns in the output of the previous model, removing the noise and hence producing results that are topologically more coherent.

Group 21: BSCS19050, BSCS19029, BSCS19019

LEARNING TO SEGMENT WHEN EXPERTS DISAGREE

1. What is STAPLE and when is it used?

Simultaneous Truth And Performance Level Estimation **(STAPLE)** is an algorithm that is used when we are doing supervised machine learning and we have contradicting ground truths as in diseases like multiple sclerosis(MS lesions) where different doctors annotate differently so it is difficult to give labels to image so here **STAPLE** is introduced which works on simple voting on each pixel. **STAPLE** will rate the accuracy of each of the radiologists and then gives weights to the annotators and then true segmentation is produced from noisy labels.

2. Why segmentation of anatomical structures is challenging even for experts?

Accurate identification of anatomical diseases like multiple sclerosis (MS) lesions MRIs is difficult even for experienced experts due to variability in lesion location, size, shape, and anatomical variability across patients. Further aggravated by differences in levels of expertise, annotations of MS lesions suffer from high annotation variations.

3. Can a model be trained without ground-truth and only noisy labels?

Yes, efforts can be made to train models without ground truth and noisy labels like as it is demonstrated in the paper that we can use a translation model of STAPLE algorithm and use CNNs(two CNNs) in which one network will measure the reliability of multiple human annotators and the other will try to estimate true segmentation labels. And both will jointly help the model in learning and predicting the unseen set of classes where no ground truths and noisy labels would be available.

Group 9 (MSDS20029 , MSDS20080) - DECAPS

How can we use CNN in MIC ? What are the limitations of CNN?

CNN is working well in the state of art in classification and object detection. But when we rotate or change the spatial context information of any object, CNN will not detect the orientation or rotation detail. Furthermore, there are many limitation, which are following below;

- CNN has positional invariance
- Lack of equivariance
- Require a lots of data
- Against HVS (human visual system).

With respect to MIC, we also want the information about patches (ROI / VOI) with their orientation, using CNN we are unable to detect these. Capsule Net is said to be viewpoint invariant that includes rotational and translational invariance.

What is a capsule network, how are they different from Convnets(CNN), how do they contribute better to MIC?

Capsule Network: consists of capsules instead of single neurons. These groups of neurons inside one capsule are responsible for defining an activity vector representing the instantiation parameters for a specific entity or part of the entity of the input image. Magnitude, which represents the probability of objects if present or not coordinate frame: encoded properties of entity (Like color, size, angle, orientation) and its angle about the orientation.

In medical imaging attending to the smallest details is critical to diagnosis of diseases. Capsule networks are good at learning these critical details; each capsule learns a specific entity as compared to ConvNet where important information is lost during max pooling and deeper layers fail to activate right neurons that capture fine grained information.

How does DECAPS improve weakly supervised learning for Medical radiology images?

DECAPS introduces Capsule heads (consisting of multiple capsules that contain multiple neurons). These capsule heads are easier to train and more detail oriented. The attention map outputs of these capsule heads can be used to make fine grained predictions of weakly labeled images for bounding box prediction. Peekaboo training uses these attention heads to do coarse predictions first then crop the ROI areas found from attention maps in raw images and use those to produce more fine grained bounding boxes. This helps the network to attend to other relevant areas in image rather than discriminative details of objects. This helps in more accurate bounding box predictions for weakly labeled medical images.

What are the improvements in DECAPS?

- They introduced Capsule heads as group of capsules & shared transformation matrix

- Introduced Inverted Routing

- Head Activation Regularization

- Peekaboo Training mechanism

Group 15 (MSCS20023) - Implicit Field Learning for Unsupervised Anomaly Detection in Medical Images

Why unsupervised learning methods are required for anomaly localization in medical images, while fully-supervised approaches are able to achieve high segmentation accuracies?

Supervised Learning methods
1) rely on the availability of large and diverse annotated datasets for training, and
2) they are specific to the anomalies annotated in the dataset and are therefore unable to generalize to previously unseen pathologies.

What do you mean by out-of-distribution anomaly detection methods? How out-of-distribution anomaly detection methods are useful in anomaly detection in medical images?

Out-of-distribution datapoint can be defined as any datapoint that was not in the training dataset and the trained model has never seen that datapoint before. In an out-of-distribution anomaly detection setting, a model usually aim at learning the distribution of healthy/normal unannotated images and at classifying as anomalies the images that differ from the learnt distribution.

Due to scarcity and imbalance of annotated datasets in the medical image domain, the out-of-distribution anomaly detection methods are very helpful.

How image reconstructions are related to anomaly detection/segmentation? Name some deep learning approaches used for anomaly detection using image reconstruction for anomaly detection?

Unsupervised Learning methods learn some intermediate representation (latent code) from a given set of training images. At inference time, the learnt representation is used to retrieve, from a given test image, a restoration, i.e. an image maximally consistent with the input one but belonging to the normal distribution. Anomalies are localized using the voxel-wise probability predicted by the model for the restored image. Some common image reconstruction based deep learning techniques are:

1. Autoencoders
2. Variational Autoencoders
3. Generative Adversarial Networks
4. Autodecoders: learn implicit field representations generated using continuous signed distance function.

Group 19: BSCS17056, BSCS18024, BSCS18037, BSCS19039
Self-Supervised Learning for MRI Reconstruction with a Parallel Network Training Framework

Q1: Why is MRI Data acquisition a slow process and how do we increase its speed while maintaining the quality of the reconstructed image?

Normally, MRI data acquisition takes a lot of time because it needs a good number of data points from which it constructs a good quality image. We speed up this process by quickly acquiring fewer data points which are known as undersampled data points and then use deep learning-based techniques to recover good quality images.

Q2: What is your understanding of k-space data in MRI?

The k-space represents the spatial frequency information in two or three dimensions of an object. The k-space is defined by the space covered by the phase and frequency encoding data. The relationship between k-space data and image data is the Fourier transformation. The data acquisition matrix contains raw data before image processing. Every point in the raw data matrix contains part of the information for the complete image.

Q3: What is SSDU (Self-Supervised Learning of Physic Guided Reconstruction Neural Networks without Fully Sampled Reference Data) in Medical Imaging and What are its Limitations?

In SSDU, we acquire MRI data in k-space and use it for image reconstruction. First of all, we divide this k-space data into two sets. One of the sets is passed

through the network and the other one used as ground truth to calculate the loss and then backpropagate it to update the parameters.

There are two important issues. First, it requires two distinct sets of data points from a single set which needs to be split with caution. When the second set does not contain enough data, the training process becomes unstable. Second, since no constraint is imposed on the unscanned data points, there is no guarantee that the final outputs are the expected high-quality images and high uncertainties exist.

Group 2: MSDS20060, MSDS20048

Q1. Why fundus images have difficulties in segmentation?

In fundus images along with influences of noises, artifacts, and illuminations, accompanied by low contrast between blood vessels and background we also have domain shift problems. That is the domain images on which we train our model are different from target domain images on which we will test our model. That makes segmentation difficult and we need to update/tweak our segmentation method to deal with these difficulties.

Q2. Fundus images are difficult to segment due to cross domain discrepancy. Which techniques can be used to solve this problem?

There are various techniques to solve this problem, some of them include.

1. Multi-label deep network and polar transformation. (M-NET) Automated segmentation using supervised CCNs.
2. Task driven generative modeling for unsupervised domain adaptation. (Transferring the input images from the target domain to the source domain, and then applying the network trained on the source domain to transferred images).
3. Cycle-GAN (Unpaired image-to-image translation using cycle-consistent adversarial networks by Zhu and fellows)
4. Patch-based output space adversarial learning (Wang et al).

However, all the above methods tried to solve this problem but failed somewhere because of domain adaptation on fundus images. Boundary and Entropy-Driven Adversarial Learning for Fundus Image Segmentation is an unsupervised domain adaptation framework. This technique aims to train images on the source domain to adapt to the target domain. BEAL uses inaccurate boundary prediction on target images and entropy driven adversarial learning to deal with

uncertainty of target domain prediction. This technique has very much better results and proved best for domain adaptation.

Q3. What is adversarial machine learning?

Adversarial machine learning is a machine learning method that aims to trick machine learning models by providing deceptive input. Hence, it includes both the generation and detection of adversarial examples, which are inputs specially created to deceive classifiers.

Types of adversarial attacks

- Poisoning attacks
  - Manipulate the **Data** before the model is used for training.
- Evasion
  - Manipulate **Model** to make incorrect prediction
- Model stealing
  - Manipulate the **Model** to learn about model and data

Group 16: Muhammad Sufyan Ashraf MSCS20072
U-Net: Convolutional Networks for Biomedical Image Segmentation

Question 1
   For what purpose U-Net is used in medical images?

Answer:
   It is used for the segmentation of images as cell segmentation is a significant task in medical images. U-Net is there to segment those cell images.

Question 2
   What is the difference between segmentation and classification? why is segmentation frequently used in medical images?

Answer:
   Segmentation models provide the exact outline of the object within an image. That is, pixel by pixel details are provided for a given object, as opposed to Classification models, where the model identifies what is in an image, and Detection models, which place a bounding box around specific objects. Medical image

segmentation has an essential role in computer-aided diagnosis systems in different applications. It divides an image into areas based on a specified description, such as segmenting body organs/tissues in the medical applications for border detection, tumor detection/segmentation, and mass detection. Because segmentation partitions the image into coherent regions, clustering procedures can be applied for segmentation by extracting the global characteristics of the image to professionally separate the ROI from the background.

Question 3

What is the Purpose of encoder and decoder in U-Net?

Answer:

An encoder in the U-Net learns to summarize (what's in the image) what kind of cells and membrane are in the image but we lose where that cell and membranes are (spatial information). The decoder in the U-Net learns spatial information.

Group 17: MSCS20030, BSCS18003

~~**What is the need of imaging in phases in liver lesion detection?**~~
~~Most of the time, the lesion of the liver is not visible only in one phase. Sometimes, it can't be distinguished from the benign lesion. So four different phases, captured with some difference of time, can help in proper diagnosis and detection of lesions.~~

How grouped SSD helps in liver lesion detection better than SSD?

Single Shot Multibox detector (SSD) is a model which passes the image through a feature extractor (VGG-16), passes the extracted feature map through multiple convolutional layers (single network with variable size layers) and does object detection with bounded boxes basing the result on best IOU. Grouped SSD is an optimized version of SSD that uses grouped convolution. Group SSD takes liver CT images of four phases. Each phase is convolved with a separate filter, by dividing the original filter. This step prevents the overfitting of a specific phase. The convolution of the four phases is then merged and an additional 1*1 convolution is applied to extract the best feature representation.

How is the loss encountered in case of SSD, when only a lesion needs to be detected?

For detection of a lesion through SSD, the loss function is the sum of confidence loss from classification head and the localization loss from the box regression head. two things need to be considered. Correct diagnosis of lesions is what we call classification, and the placement of the bounding box as close to ground truth is what we call localization. Localization loss is computed by calculating the difference between both boxes (predicted and ground truth) by passing it through a function, which in this case is smooth L1 loss. The localization is only computed of boxes that are labeled to be positive boxes and not negative ones. The confidence loss is the loss of making a class prediction. For every positive match prediction, we penalize the loss according to the confidence score of the corresponding class. For negative match predictions, we penalize the loss according to the confidence score of the class "0": class "0" classifies no object is detected.

What is difference between one stage detectors and two stage detectors?

- **The one stage detectors treat object detection as a simple regression problem by taking an input image and learning the class probabilities and bounding box coordinates. Such models reach lower accuracy rates, but are much faster than two-stage object detectors. E.g. SSD, YOLO etc**
- **Two stage detectors use a Region Proposal Network to generate regions of interest in the first stage and send the region proposals down the pipeline for object classification and bounding-box regression. Such models reach the highest accuracy rates, but are typically slower. E.g. R-CNN and Mask R-CNN etc.**

Group 10: Khuzaima Shahid MSDS20039, Abdullah Aziz MSDS20052

Q: What is meant by semantic segmentation and how is it important in medical imaging?
Semantic segmentation means labeling all pixels in which 1 would mean presence of an object or class and 0 would mean absence. It does not differentiate between objects of the same class. For example, all pixels containing a dog will be labeled as 1 and all other pixels as 0. In Medical domain, the majority of the tasks are related to semantic segmentation in which imaging technologies such as X-rays, MRIs and CT scans are used to segment tumors, lesions and masses.

Q: What is meant by Image super resolution and how is it important in medical imaging?
Image super resolution is a sub-problem of image to image translation in which a lower resolution image is upsampled through a model to produce high quality images. This is extremely important in the medical field due to multiple reasons. Firstly, it can be used to remove shadows or noise from a medical image which would result in better diagnosis.

Secondly, most of the medical imaging equipment is extremely expensive. However, image super resolution can be used to make medical images taken from cheap equipment as close as possible to one taken from an expensive machine.\ which would benefit accessibility and elad to better diagnosis.

Q: How does Bio-Net help in the field of medical image computing?

 Autoencoders are used widely in medical image computing ranging from segmentation to image super resolution to anomaly detection. Bio-Net has proposed a few novelties to autoencoder architecture. Firstly, they have added forward skip connections as well as backward skip connections in a UNET architecture. Secondly, they have a recursive inference mechanism that reuses the same encoder and decoder blocks in a recursive manner with shared parameters. In short, they have shown gains like cascaded autoencoders while retaining the parameter space of a single autoencoder. These novelties can be applied to any autoencoder to boost performance while retaining the same parameter space.

 Group 11

KiU-Net: Towards Accurate Segmentation of Biomedical Images using Over-complete Representations

MSDS20044 (Hafiz Muhammad Arslan)
MSDS20015 (Hira Ejaz)

1.   How to learn fine detail like edges in segmentation

In segmentation fine detail or low level features can be learned through either skip connections or through over-complete architecture where the intermediate layers are of high dimensions in an encoder-decoder style network. This is because such architecture learns finer low-level features due to the decreasing size of the receptive field even as we go deeper in the encoder network.

Skip Connections (or Shortcut Connections) as the name suggests skips some of the layers in the neural network and feeds the output of one layer as the input to the next layers.

2.   How (Why) to augment two network i.e. U-Net and Kite-Net

Because there are two networks parallel and work simultaneously and give low level and high level features we need to augment these two features. We can augment it by just

adding but this is not an optimal way to do that so they introduce a block Cross Residual Fusion Block to augment feature of both networks i.e. Ki-Net and U-Net at each layer. By using cross residual fusion block (CRFB) block you need to extract complementary features from both network branches and forwards to both of them respectively. Specifically, the CRFB consists of residual connections, followed by a set of conv layers. In order to combine the feature maps from the two networks F_U^i (U-Net) and F^i_Ki (Ki-Net) after the ith block, cross-residual features R_U^i and R_Ki^i are first estimated through a set of conv layers. These cross-residual feature are then added to the original features F_U^i (U-Net) and F_Ki^i to obtain the complementary features F_U^i and F^Ki i , i.e, $F^Ui = FUi + RKi i$ and $F^Ki i = FKi i + RUi$. This strategy is more effective compared to simple feature fusion schemes like addition or concatenation.

3.   What is under complete and overcomplete architecture

Under-Complete Convolutional autoencoder are those where the dimensionality of data is reduced near the bottleneck. In this way the size of receptive field increase and filters only learn high level details like objects, blob etc.

Over-Complete Convolution architectures are those where the data is projected onto a higher dimension in the intermediate layers. In this way the size of the receptive field decreases as we go deeper in the layers and filters are able to learn low level details like edges.

Group 18: Semi-supervised medical image classification with relation-driven self ensembling model

Group members: MSCS20014(KINZA FAYYAZ), MSCS20012(HIJAB SEHAR)

Question#1: What is the difference between multi-class classification and multi-label classification in medical imaging?

Answer: Multiclass classification makes the assumption that each sample is assigned to one and only one label. For example, in our research paper, they are using the skin lesion dataset that is the multi-class problem. We classify the 7 types of skin lesion which may be melanoma, Melanocytic nevus, Basal cell carcinoma, Actinic keratosis, Benign keratosis, Dermatofibroma, Vascular lesion. But a person can have only one type of skin lesion that is mentioned above.

Multi-label classification assigns to each sample a set of target labels. This can be thought of as predicting properties of a data point that are not mutually exclusive. For example, chestxray14 is a multi-label classification used in our research paper. A patient can

have multiple diseases like pneumonia, Mass, Nodule, Edema, etc.

Question#2: What is Sample relation consistency and How semi-supervised method of Sample Relation consistency (SRC) is exploiting unlabeled data?

Answer: Our proposed SRC(sample relation consistency) is a semi-supervised learning framework that explicitly enhances the consistency of the intrinsic relation among different samples. SRC encourages the consistency of the structured relation (similarity) among different samples, highlighting that samples with high similarity should also be highly related after adding perturbations. It is a consistency-based method that exploits the unlabeled data by encouraging the prediction consistency of given input under perturbations and leverages a self-ensembling model to produce high-quality consistency targets for the unlabeled data. Our framework regularizes this relation structure to be stable between the teacher model and the student model under different perturbations to extract richer semantic information from the unlabeled data and uses all the useful information of unlabeled data to train the model.

Question#3: What are perturbations in medical images? What are their benefits when we add perturbation in input images of chest x-ray and skin lesions?

Answer: Perturbation is just noise. In our semi-supervised method, the Sample Relation consistency (SRC) paradigm requires the consistency of the intrinsic (more detailed) relation of data under different perturbations. Two types of perturbations were added to chest X-ray and skin lesian input images during training, including 1) Random rotation, translation, horizontal flips to input samples. 2) We added a dropout layer before the last pooling layer in the employed Dense Net, with a dropout rate of 0.2. The magnitude of these perturbations were chosen based on empirical observations in preliminary experiments.designing a more effective perturbation scheme is able to improve the performance of consistency-based methods. The benefit of perturbation is that the model can be trained for unknown images that can be used as augmented data.

Group 4

Source-Relaxed Domain Adaptation for Image Segmentation

Muhammad Ahsan Akhtar (MSDS20078)
Muhammad Taimur Adil (MSDS19040)

Q1:What are the techniques for Domain Adaptation?

There are three techniques for Domain Adaptation

1. Divergence based Domain Adaptation

2. Adversarial based Domain Adaptation

3. Reconstruction based Domain Adaptation

Q2:Domain Adaptation is a technique in which we learn a model on a training dataset and then apply it on a slightly different dataset which is rare as well. In case of domain adaptation where we don't have access to the source dataset, why do we add KL (Kullback-Leibler) divergence in the loss function?

In domain adaptation where we have both source and target dataset our loss function depends upon standard supervised loss. If we use this loss for the current case where we don't have access to ground truth, then our loss function will be biased towards the class that is in majority in the training dataset. To overcome this issue, we first synthesize our ground truth with the help of an auxiliary network then apply KL divergence on the ground truth and testing data. KL divergence gives us the measure about how similar distributions are. Adding KL divergence with the standard supervised loss gives us a good loss function that is not biased towards the majority class in the testing dataset.

Q3:Entropy map is the calculation of approximation of interested region presence. How is it created in case of Domain adaptation where we don't have access to source data?

Entropy map is created with the help of segmentator and two fully connected layers , acting as soft segmentation, applied to the training data.

Group 14 : Hardnet

- **MSDS20031 Maham Ilyas**
- **MSDS20081 Sidrah Jibran**

Q1. What are skip connections and what are the benefits?

It skips some of the layers in the neural network and feeds the output of one layer as the input to the next layers.
Skip Connections were introduced to solve different problems in different architectures. In the case of ResNets, skip connections solved the degradation problem that we addressed earlier whereas, in the case of DenseNets, it ensured feature reusability.

Q2. What is inference time and its role in Medical imaging?

Inference time is the time it takes for a deep learning algorithm to apply the trained neural network model to new data. This time can be affected by the complexity of the network, the number of layers, and the number of neurons in each layer. Generally, inference time increases with network size and complexity. It is important in Medical imaging as we are trying to automate the process and make it less time-consuming.

Q3. How can the Receptive field be used as a decoder in Medical imaging to capture deep features ?
It can strengthen the deep features learned from a lightweight CNN backbone. By using multi-branch with different kernel size convolution and dilated convolution layers, it generates features with the different receptive fields.

Group : 08

1: MSDS20074 Abdul Hanan

2: MSDS20083 Abdul Rehman

Paper Title: CGC-Net: Cell Graph Convolutional Network for Grading of Colorectal Cancer Histology Images.

Link : https://arxiv.org/pdf/1909.01068.pdf

Slides Link :
https://docs.google.com/presentation/d/1tbFuC4k5zW05jhf6eSoo9alfBcfrpddvpLxW6Axh4Qc/edit?usp=sharing

Q#1: Why can we use GCN instead of CNN for Whole slide images(WSI)?

The major difference between CNNs and GNNs is that CNNs are specially built to operate on regular (Euclidean) structured data, while GNNs is the generalized version of CNNs where the numbers of node connections vary, and the nodes are unordered. So in WSI that converts each large histology image into a graph, where each node is represented by a nucleus within the original image and cellular interactions are denoted as edges between these nodes according to node similarity. The GCN utilizes nuclear appearance features in addition to the spatial location of nodes to further boost the performance of the algorithm.

Q#2: Why do we combine Hierarchical features for graph-level classification?

To utilize a max operation for the node embeddings at each stage to get a fixed-size representation. Then the concatenation of multi-level representations is fed into the linear layer to get the prediction for classification.

Q#3 Why do we use clustering instead of pooling in GCN?

The node features contain the local contextual information. However, the flat structure remains a drawback because the hierarchical structure is lost when using global mean/max pooling to get the graph-level prediction. Therefore, the clustering operation is necessary to extract more abstract features for hierarchical representation.

Q#4 On which data GCN can be applied in medical Images?

GCNs can take a certain length of nucleus fingerprints as input and generate predicted structures of the nucleus. CGC-Net of Graph Convolutional network which helps to create nucleus structures with various features in it. In this way, it allows scientists to identify the stage and grade of cancer.

Group : 01

1: MSDS20051 Bilal Ayub

Paper Title: DoubleU-Net: A Deep Convolutional Neural Network for Medical Image Segmentation

Q1. Why does DoubleU-Net i.e. combination of Two U-Net architectures perform better than single UNet?

Ans. Output feature map from the first U-Network is concatenated with the Input Image and given to the second U-Net as Input gives our model a second chance to improve the output mask by focusing on the region which is common between the first network output mask and the image. Moreover, better pooling techniques i.e., ASPP and squeeze-and-excite block makes the performance of Double-UNet better than single UNet.

Q2. What kind of pooling layer can be added to the deep neural network that can help in preserving multi-scale contextual information for a medical image?

Ans. We can use ASPP i.e. Atrous Spatial Pyramid Pooling layer. It resamples the feature layer at multiple rates prior to convolution. The resultant output is a pyramid of feature maps each describing features at different scales in the same input image. These feature maps are concatenated and this allows us to easily and explicitly control spatial resolution by aggregating multi-scale contextual information in our network.

Q3. How does Medical Images make Image Segmentation tasks challenging?

Ans. In medical images, there are images belonging to different modalities, shapes and sizes. Moreover, there is unavailability of a large number of annotated datasets, there is lack of high-quality images, lack of standard segmentation protocols. All of this makes the segmentation task challenging for Medical Images.