19EE10039
MANSI UNIYAL

Q1 In what way is image segmentation useful for Remote Sensing? Explain with examples.
Explain in detail how we may use Deep Learning to get accurate rainfall estimates from two different remote sensing sources, like a ground-based and a space-borne radar.

Image segmentation is the task of classifying each pixel into the labeled category of classes. Image segmentation is useful for Remote Sensing in many ways like:

1. Tracking clouds to help predict the weather or watching erupting volcanoes, and help watch for dust storms. With time analysis of weather conditions changing. Even large forest fires can be mapped from space, allowing rangers to see a much larger area than from the ground.
2. Analyzing the growth of a city and changes in farmland or forests over several years or decades. Urban Planning: building and road mapping.
3. Check on Deforestation over time, crop production mapping, land cover analysis, understanding the spatial distribution of water use anticipating the likelihood of water scarcity, and as a result, food insecurity.
4. Discovery and mapping of the rugged topography of the ocean floor (e.g., huge mountain ranges, deep canyons, and the "magnetic striping" on the ocean floor). These can enable understanding of geographical structures and evolution over time. Even mountain regions, hilly areas and plain cultivated lands can be analyzed and used either for civilization or extraction of minerals and other usages.

Additionally, the various examples of usage of image segmentation apart from remote sensing:

- Autonomous vehicles: We need to equip cars with the necessary perception to understand their environment so that self-driving cars can safely integrate into our existing roads.
- Medical image diagnostics: Machines can augment analysis performed by radiologists, greatly reducing the time required to run diagnostic tests. Like segmenting brain tumors.

Image segmentation is used to get accurate rainfall estimates from two different remote sensing sources like ground radar and satellite images.
Training of a simple MLP model to ground radar images to true values that is growth truth (rain gauges) and other MLP to map space borne images to ground radar images. That is X1, ground radar images, is mapped to Y, precipitation values, and X2, space borne images, to X1. Multiple preprocessing is done like vertical and horizontal alignment of both sensors.

$$\mathbf{y_1} = f(\mathbf{w}_1\mathbf{X} + \mathbf{b}_1)$$
$$\vdots$$
$$\mathbf{y_n} = f(\mathbf{w}_n\mathbf{y_{n-1}} + \mathbf{b}_n)$$
$$\mathbf{Z} = f(\mathbf{w}_{n+1}\mathbf{y_n} + \mathbf{b}_{n+1})$$

Machine learning-based approach support in rainfall algorithm development for the dual-frequency precipitation radar. Where the dual-frequency measurements can be suited to the MLP framework in order to better account for the precipitation microphysical variability. In addition, we can incorporate additional input features such as numerical weather prediction model results.

Q2 In what ways can we identify different predictors of Indian monsoon rainfall using Machine Learning? Explain with clear formulations.

We need food to live and the main source of food comes from agriculture. Rainfall acts as one of the prime factors for healthy agriculture. Rainfall prediction is thus an important task.

Several deep learning methods are being used to map features like climatic variables: air temperature, rainfall, and sea level pressure. But as rainfall depends on various factors, learning such mappings is difficult.

With features like temperature, humidity, height from sea level along with raw input data, any regression-based Machine Learning technique, such as the ensemble regression tree model, or SVM may be used to forecast monsoons using these few important detected climate indicators.

A raw satellite image can even be given as input to deep learning models like stacked autoencoders that can be utilized to get the key rainfall factors. Encoder encodes the raw input to latent representation and the decoder decodes these latent variables to Pixel-pixel wise precipitation values. Other models like SVM, MLP can also be used for the same.

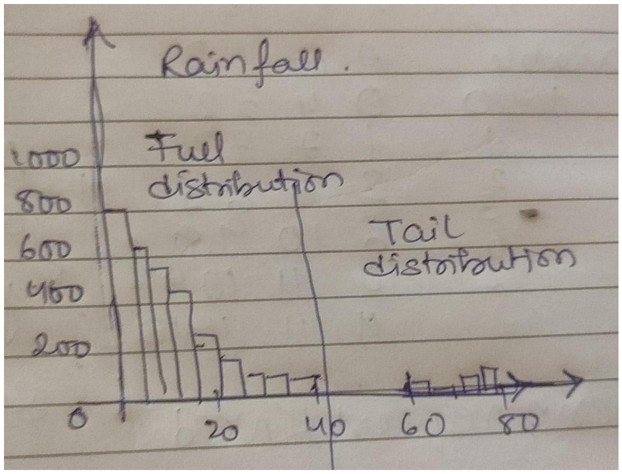
Finally, improvements in the model of deep learning autoencoders and extracting useful latent features can help enhance rainfall prediction quality.

Q3 Explain with an example, why we need Extreme Value distributions like GEV, instead of using usual distributions for climatic variables?

Dealing with the characteristics of the tail of the distribution is the study of Extreme value statistics. It doesn't focus on the main distribution. Thus these Extreme value distributions are referred to as the limiting distributions for the maximum or minimum of the group of independent random variables.

Let, F(x) be a continuous initial distribution and have an inverse. Then, CDF of GEV distribution:

$$F_{m\_n}(x) = P(m\_n<=x) = \exp\{ -[1+\xi((x-\mu)/\sigma)]^{1/\xi} \}$$



Hence the Full distribution can be fitted to any other exponential or gamma distribution but the Tail distribution is very different. Thus the need of a special type of distribution like GEV is present to understand the same.
Techniques like: block maxima approach or peak over the threshold can be used to fit.
For calculating the extreme value,
By using normal gamma distribution in the ft collins case,
P(annual_max >6.18) = 6.86×10^{-9} years
By using GEV distribution on block maxima will give the approximately correct analysis,
P(annual_max > 6.18) = 121 years

Thus Extreme Value distributions like GEV are needed, instead of using usual distributions for climatic variables.

Q4 Explain how data assimilation works using a Kalman Filter.

Due to model and include uncertainties the model output will not always correspond to reality completely that's where data and simulation comes in handy. This is called Data Assimilation. Several methods can be used for data assimilation:

- Kalman filter
- Ensemble Kalman filter
- Deterministic Ensemble Kalman filter
- Partial deterministic ensemble Kalman filter

Kalman filter built on the principle of Bayesian inference. It is used to estimate a probability distribution for all model variables. It requires the model to be linear and observations to be stochastic and Gaussian distributed. It uses Bayes theorem to update a distribution probability based on a prior distribution, a likelihood distribution and the strength or uncertainties.

Using Bayesian inference the updated model state moves closer to what's the likelihood probability distribution with every update. If the standard deviation for the likelihood had been less than the prior meaning, then the observation should be trusted more the update would show the impact of the uncertainties and probability distributions.

By using the patient inference to update the distribution probability of the model variables, the Kalman filter can estimate the best model state and its error covariance.