MD MESBAHUR RAHMAN

Japan, Tokyo, Adachi-ku, Takenotsuka 1-6-7-404
+817075265103 • md.mesbahur.rahman@utexas.edu
Website: https://mmrahman-utexas.github.io/
LinkedIn: https://www.linkedin.com/in/md-mesbahur-rahman-1870b514b/
GitHub: https://github.com/mmrahman-utexas

Work Experience

Computer Vision Engineer, DeepX Inc., Tokyo, Japan **February 2022 – Present**

- Computer Vision module development, data analysis, consulting and administrative duties

AI Engineer, Hiperdyne Corporation, Tokyo, Japan **November 2018 – January 2022**

- Solved client problems using deep learning and machine learning approaches

Teaching Experience

Lecturer, Uttara University, Dhaka, Bangladesh **September 2018 – November 2018**

- Taught academic courses at the department of Electrical and Electronic Engineering

Education

The University of Texas at Austin, Texas, United States

- Master of Science in Computer Science
- Fall 2020 – Fall 2023
- Coursework: Advanced Linear Algebra , Machine Learning, Deep Learning, Reinforcement Learning, Natural Language Processing, Optimization, Online Learning and Optimization, Algorithms: Techniques and Theory, Advanced Operating Systems, Case Studies in Machine Learning

Bangladesh University of Engineering and Technology (BUET), Dhaka, Bangladesh

- Bachelor of Science in Electrical and Electronic Engineering
- February 2013- September 2017
- CGPA: 3.69 on a scale of 4.00
- Thesis Title: 'Numerical Simulation and Analysis of Room Temperature Plasmonic Laser'
- Paper accepted at 2018 7th IEEE International Conference on Photonics (ICP), Malaysia

Dhaka College, Dhaka, Bangladesh

- Higher Secondary School Certificate, Science
- April 2010 – March 2012
- GPA: 5.00 on a scale of 5.00

Standardized Test Scores

- **GRE:** 318 (Verbal Reasoning:154, Quantitative Reasoning:164, Analytical Writing:3.5)
- **TOEFL (Expired):** 110 (Reading: 30, Listening: 30, Speaking: 22, Writing: 28)

Technical Courses and Training

Udacity, Inc., California, United States

- Computer Vision Nanodegree ▪ Status: Graduated on May 29, 2020
- Deep Learning Nanodegree ▪ Status: Graduated on May 13, 2020

Hiperdyne Corporation, Overseas Division, Dhaka, Bangladesh

- Extensive Artificial Intelligence Training Course ▪ September 2017- January 2018
- Course Title: 'Traditional Machine Learning and Deep Learning with Python'

Technical and Relevant Skills

- **Programming Languages:** C, C++, Python, MATLAB, SQL, Assembly
- **Programming Tools and Frameworks:** Scikit-Learn, Keras, TensorFlow, PyTorch, LibTorch, HuggingFace, OpenCV, NLTK, TextBlob, Pandas, NumPy, SciPy, Matplotlib, Seaborn, PCL, Open3D, Eigen etc.
- **Industry Knowledge:** Object-Oriented Programming, Machine Learning, Deep Learning, Data Science, Computer Vision, NLP, Reinforcement Learning, MLOps, etc.
- **Algorithms:** CNN, VGG, Inception, Xception, Resnet, FCN, SSD, Retinanet, Transformer
- **Software Development:** Git, Docker, GitHub Actions, CI/CD, DVC, MLflow, ONNX, TensorRT, Conda, CMake, PyCharm, Visual Studio Code, Visual Studio, ROS2, Rviz2, etc.
- **OS & Cloud Platform:** Windows, Linux, macOS, AWS, Google Colab
- **Academic:** LaTeX, Overleaf, Papers With Code, etc.
- **Language Skills:** Bangla, English, Hindi, Japanese (N5 Level), Korean (Basic)

Project Experiences

Point Cloud Segmentation: Training DeltaConv for Segmenting Dump-bed

- Generated 3D Point Cloud simulation data consisting of labeled terrain and dump-bed points of dump-trucks using proprietary simulation tool of DeepX, Inc. incorporating data augmentation. Trained and tuned DeltaConv model using the implementation provided by the authors of this paper for segmenting the dump-bed points of an excavator seen by the scan of a Lidar sensor.

Refactor C++ ROS2 Nodes to Create a ROS Independent C++ Library Package

- Designed steps for extracting reusable computer vision modules out of a C++ codebase written to be a set of ROS2 nodes to make them ROS independent. Transferred the code from computer vision nodes written in C++ to header files (.hpp) refactoring the code to be in form of ROS independent classes. Designed ROS independent data structures to communicate with main ROS2 node that imports the C++ header files and uses the reusable computer vision classes.

Design and Develop Linear In-painting algorithm for heightmap for DeepX, Inc.:

- Designed and developed an heightmap in-painting algorithm by identifying the convex-hull of the missing areas and then extending the edge of the missing area linearly along the direction of image gradients and vertically along the direction of a tunable slope angle up to a user defined height. Used image processing functions and tools from OpenCV,

Scikit-Image and NumPy to implement the linear in-painting algorithm for filing in the missing pixel in a heightmap generated from a PointCloud.

Design and Develop PointCloud Filter based on geometry for DeepX, Inc.:

- Designed and developed a ROS2 based pass-through range filter in C++ for removing points of a pyramid shape building from a PointCloud using a geometry-based filtering. Generated simulation PointCloud test data and deployed the filter on site after testing its performance on real PointCloud test data.

Design and Implement Rule Based Anomaly Detection Algorithms for DeepX, Inc.:

- Created real image dataset of industrial sample products by creating scratch and dents defect on them and taking images using industrial grade camera, lens and lighting equipment. Designed and implemented rule-based computer vision algorithms for scratch and dent defect detection in images of industrial sample products. Some of the notable implemented rule-based AD methods are “Adaptive Thresholding”, “Edge Detection”, “Morphology”, “FFT Bandpass filtering”, “Texture Entropy Analysis”, “Texture Laws Filtering” etc. which were implemented using Halcon and OpenCV frameworks. These algorithms were developed as part of the project (ongoing) of “Automation of Design Process of Visual Inspection System (Imaging Conditions and Inspection Algorithms)”.

Develop ROS2 application for Communication Between Python and C++ nodes:

- Created a ROS2 application consisting of a python publisher node and a C++ publisher node and a C++ subscriber/publisher node for a warmup project at DeepX, Inc. Developed and built the ROS2 package successfully that reads a PNG RGB image file using a python node and a PCD point cloud file using a C++ node and publish them to two topics and publish the image size and point cloud size using a C++ node by subscribing to the two topics.

Image Classification and Object Detection of Construction Image for So-net:

- Annotated five categories of optical fiber construction objects using “labelImg”, an image annotation tool and modified the annotation output format to meet the specific need of the project. Then Fine-tuned a custom yolov4 model with pre-trained weight set to detect five categories of objects in the input images and then built a prototype python application for cropping into the input images for the detected five categories of objects or classifying the object images using the bounding box predicted by the yolov4 algorithm and organized the detected object images into five corresponding folders as per requirement of So-net (Sony Network Communications Inc.).

Object Detection – Determining Object Location and Size Implementing CenterNet:

- Built an object detector from scratch inspired from this paper, by training a Fully Convolutional Network for predicting heatmap for each class and then extracting the center/peak by a peak extractor function also predicted the size of the object in two separate output channels. Jointly optimized combination of focal loss and L1 loss for training model while the focal loss for predicting the class heatmap and the L1 loss for predicting object size and achieved the target performance metric set by course teacher.

Semantic Segmentation – Image Segmentation Using Fully Convolutional Network:

- Defined a Fully Convolutional Network with up-convolution layer, residual and skip connection in PyTorch and trained it on the dense-labeled “SuperTuxCart” dataset to

predict segmentation mask for correct classes and achieved 100% score in all target performance metrices set by the course instructor.

Image Anomaly – Unsupervised Anomaly Detection Using Convolutional Autoencoder:

- Built an unsupervised dataset from a supervised labeled dataset of MNIST dataset by removing its labels. Then defined a Convolutional Autoencoder network in PyTorch and trained it on the unsupervised dataset and allowed the network to learn to reconstruct the training images containing regular images with a small percentage of anomaly images. Then during inference, calculated a reconstruction error (MSE) threshold based on a given percent quantile and declared an input image as an anomaly for who's output image reconstruction error is above the preset threshold.

Image Anomaly – Anomaly Detection & Localization by Fine-tuning VGG19:

- Partially fine-tuned a pre-trained VGG19 model on a data-augmented supervised training dataset containing images of walls, containing cracks and no cracks, to obtain a test accuracy of 92% for detecting crack in the test images and localized the cracks afterwards using activation maps from the last convolution layer of the VGG19 model and by projecting the activation maps after reshaping it by linear-up sampling and taking dot product with the corresponding class weights of the last fully connected layer of the modified 'classifier' module of the VGG19 model to figure out the activation for the predicted class.

SLAM – Implementing Landmark Detection & Robot Tracking through Graph SLAM:

- Implemented SLAM (Simultaneous Localization and Mapping) specifically Graph-SLAM for a general 2-dimensional world. Combined robot sensor measurements and movement model to create a map of an environment from only sensor and motion data gathered by a robot, over time by calculating Omega and Xi constraint matrices incorporating elements of uncertainty, probability and linear algebra.

Caption Generation – Image Caption Generation using CNN-LSTM Encoder-Decoder:

- Pre-processed the images in the MS COCO Dataset using PyTorch Transforms and converted the captions in the training set into sequence of integers using BOW vocabulary dictionary with a vocabulary threshold of 5. Defined and trained a CNN encoder and a LSTM Decoder on top of a time distributed embedding layer by using pre-trained RESNET50 model as a feature extractor to encode an input image into a fixed embed sized vector and then used LSTM decoder to generate captions from the output embedding vector of the CNN encoder. Configurations of the data pre-processing and CNN encoder and LSTM decoder were inspired from this paper. Then inference was done on the 'test' portion of the MS COCO dataset.

Landmark Detection – Facial Keypoint Detection using CNN, Haar Cascade Classifier:

- Defined and trained a CNN on facial keypoint dataset from "YouTube Faces Dataset" using custom transformation in PyTorch to perform regression task to predict the location of 68 facial keypoints as inspired from this paper. During inference detected all the faces in an image using OpenCV's pre-trained Haar Cascade classifiers and predicted the location of 68 facial keypoints on those detected faces using our trained CNN network.

Image Segmentation – Unsupervised Image Segmentation Using K-means Clustering:

- Implemented K-means clustering based on RGB values of input color image using OpenCV's Kmeans function by defining k=3 and criteria for convergence combined of

maximum iteration number and epsilon and visualized the different segments of the input image by using clustered pixels as a mask.

Design and Building Python Library and Application for Music Information Retrieval for Sony CSL:

- Designed and built a python library to be imported from another python script in modular fashion for music information retrieval (e.g. beat, chord, genre, mood, bpm, segment prediction etc.) from input wav or mp3 song file using machine learning and Deep learning models for Sony CSL (Sony Computer Science Laboratory). Also built a command line python application to show the usage of the library. Currently in charge of adding new features to the library and multiple OS support to the library (Ongoing).

Sequence Classification – Error Pattern Detection in Traditional Beat Tracking for Sony CSL:

- Implemented sequence classification using hierarchical attention for an LSTM sequence Autoencoder on LibROSA extracted features to classify error pattern in beat tracking done by a proprietary music analysis tool for Sony CSL (Sony Computer Science Laboratory) at Hiperdyne Corporation.

Music Information Retrieval - Music Genre/Mood Classification for Sony CSL:

- Trained various deep learning model and machine learning model on a structured dataset by extracting numerous music related features from mp3 files using a proprietary music analysis tool and achieved beyond state of the art for predicting multi-label genre/mood of a song for Sony CSL (Sony Computer Science Laboratory) at Hiperdyne Corporation.

Model Deployment – Deploying a Sentiment Analysis Model Using AWS SageMaker:

- Constructed an RNN (LSTM) on top of an embedding layer and trained it on the BOW encoded IMDB dataset containing movie review for detecting sentiment of a movie review. Created the training job, model and deployed the model on AWS SageMaker service. Gave access of the deployed model to a simple web app that can take input review as text from a user and give the sentiment of the review as output by creating a Lambda function and API Gateway and giving the web app access to the API Gateway.

Image Transformation – Unpaired Image to Image Translation Using CycleGAN:

- Built a CycleGAN and trained on the unlabeled and unpaired dataset images of "summer" and "winter" from "Yosemite National Park" dataset to transform the images from the "summer" domain to "winter" domain and vice versa. Defined two sets of Discriminator and Generator, each for transforming images from summer to winter and vice versa. Trained the two Discriminator separately by optimizing respective discriminator losses and the two Generator together by optimizing addition of two generator losses and two cycle consistency losses and visualized the output images, transformed by the trained Generator networks, across the two domains.

Image Generation – Generating Realistic House Numbers Image using DCGAN:

- Defined Deep Convolutional Discriminator and Generator network using Conv, transpose Conv and batch Norm layer. Trained the discriminator network optimizing combined discriminator loss on real and fake images using true labels and trained the generator network optimizing discriminator loss on fake images using flipped labels to train the entire GAN on Google SVHN dataset to generate realistic looking house number images.

Text Generation – TV Script Generation using NLP:

- Built a system for pre-processing text in the TV script of "Seinfeld" dataset and constructed TV script generator using a LSTM network with word embedding in PyTorch and trained the model on the "Seinfeld" dataset with hyper parameter tuning to achieve a "Categorical Cross Entropy" loss of 3.24 for vocabulary size of 21388 in training set for script generation.

Word2Vec – Training and Visualization of a Word Embedding Space:

- Built a word embedding based on "text8 dataset". Trained the embedding weight matrix in 'skip gram' fashion and did sub-sampling the common word based on their frequency. Also visualized the word embedding space after reducing dimensionality using T-SNE method.

Object Detection and Classification – Dog Breed Classifier:

- Built a pipeline to first detect a human or dog in a user provided image using OpenCV's implementation of Haar-cascade classifiers and VGG-16. Then predicted the breed of the dog or the most resembling dog breed of the human face by a CNN model. Implemented transfer learning in PyTorch by using VGG-19 model as a feature extractor and training a data-set specific classifier to achieve 86% test accuracy for detecting correct dog breed based on a dataset consisting of 133 dog breeds.

Image Composition – Style Transfer Using Pre-trained VGG19 Model:

- Extracted content representation of the input content image from the activation map 'Conv4_2' layer of a pre-trained VGG19 model and style representation of the input style image from calculation of Gram Matrix from the activation maps of the 'Conv1_1', 'Conv2_1', 'Conv3_1', 'Conv4_1', 'Conv5_1' layers of the pre-trained VGG19 model as mentioned in this paper and optimized combined content and style loss between the content and style representation of the target output image and those of the content image and style image to transfer the style from 'input style' image to 'input content' image.

Image Translation – Denoising MNIST Dataset using Convolutional Autoencoder:

- Built and trained a Convolutional Autoencoder network on an artificially noised MNIST dataset to denoise the input image by performing an image translation task using the convolutional autoencoder and obtained a training pix2pix MSE loss of 0.46 and visualized the output denoised images.

Regression Problem – Predicting Bike-Sharing Patterns:

- Built a neural network from scratch using NumPy and predicted how many bi-cycles a bike-sharing company would need on a given hour of a day given some features of historical fashion (e.g., time, temperature, humidity, holidays and weekdays information etc.). Tuned epoch number, learning rate, hidden layer and hidden layer's node number of the training to achieve a training MSE loss of 0.09 and validation MSE loss of 0.18.

Multi-label Text Classification – Predicting Lyric Theme from Song Lyrics:

- Implemented different machine learning and deep learning architecture to solve a multi-label classification problem for predicting various lyric-theme of a song lyrics in text form for a client company at Hiperdyne Corporation. Used frameworks like Scikit-Learn and Keras to train on features extracted using TF-IDF and Word2Vec to achieve best in class multi-label F1 score lyric-theme prediction model.

Sentiment Analysis – Detecting Sentiment from Comments in Social Media:

- Performed a data analysis and data cleaning operation on a dataset consists of English and Bangla (Both in Bangla and English phonetic alphabet) comments from social media. Pursued Hype-parameter tuning implementing GridsearchCV of different machine learning model (e.g., Multinomial Naïve Bayes, Logistic Regression etc.) to achieve satisfactory accuracy for detecting Sentiment in social media comments.

Binary Classification Problem- Titanic: Machine Learning from Disaster:

- Built various machine learning model (e.g., KNN, Random Forrest, SVM, Logistic Regression etc.) and trained on a data set consists of numerous passenger information after doing data-preprocessing and cleaning operation to predict whether the passenger survived or not.

Achievements

- Got selected for Dean's List Scholarship 2015 in the faculty of EEE, BUET
- Was Finalist in 4th National Chemistry Olympiad, 2011, Bangladesh
- Was selected Prime Bank Foundation Scholarship, 2010

Reference

Dr. Md. Zahurul Islam
Professor
Department of EEE, BUET
Email:
mdzahurulislam@eee.buet.ac.bd

Abdus Salam Azad
Assistant Professor,
Department of CSE, BUET
Email: azadsalam@cse.buet.ac.bd