Simmi Mourya

https://simmimourya.github.io/

Email : simmim@seas.upenn.edu
Philadelphia, PA 19104

Education

- **University of Pennsylvania** — Philadelphia, PA
  Master of Science in Computer and Information Science, GPA: ***3.67/4.0*** — ***Graduating May 2021***
- **Coursework**: Advanced Machine Perception, Computer Vision, Internet and Web Systems, Computational Linguistics, Machine Learning, Independent Research: *Video Object Segmentation* (advised by **Prof. Jianbo Shi**)
- **Cluster Innovation Center, University of Delhi** — Delhi, India
  Bachelor of Technology in Information Technology — *Aug. 2013 – July 2017*

Skills

- **Research**: Python, PyTorch, Keras, Scikit-Learn, Numpy, Pandas, Caffe, Cython, Python/C API
  Software: Java, Apache Spark, Apache Storm, Apachebench, Oracle BDB, PHP, HTML/CSS, SQL, Nose, Jenkins, ArcGIS

Experience

- **ESRI** — Delhi, India
  Data Scientist — *May 2019 - July 2019*
  - **ArcGIS Python API**: Developed framework for **Multispectral support for Pixel classification** in **ArcGIS** Python API. Developed Pyramid scene parsing backbone support of **object segmentation** for the API using PyTorch and FastAI.
  - **Spatial Dataframes**: Optimized validation checks in arcgis.geometry package using pre-compiled Cython binaries. Now processes **0.1 million entries in less than 2 ms**, which earlier took **45-55 ms**.
- **IIIT Delhi** — New Delhi, India
  Research Associate — *Feb 2018 - March 2019*
  - **Article**: Mourya, S., Kant, S., Kumar, P., Gupta, A. and Gupta, R., 2018. LeukoNet: **DCT-based CNN** architecture for the **classification** of normal versus Leukemic blasts in B-ALL Cancer.
  - **Accepted Challenge**: **Classification of Normal versus Malignant Cells** in B-ALL White Blood Cancer Microscopic Images, challenge selected at IEEE ISBI '19, Venice, Italy.
  - **R&D**: **Designed and deployed** (at AIIMS hospital) LeukoAnalyzer - Fuses Discrete Cosine Transform **(DCT)** domain features extracted via CNN with the Optical Density **(OD)** space features for **detection of cancerous white blood cells** from blood slides. Also developed LeukoGAN: A **Dual representative adversarial** network based on **U-Net** inspired **ACGAN** to generate synthetic B-ALL Cancer data.
- **Predible Health** — Bangalore, India
  Deep Learning Developer — *August 2017 - December 2017*
  - **Development**: Developed U-Net based framework for **Lung nodule segmentation** from 3D CT scans (LIDC-IDRI dataset) using PyTorch and Python Scientific Stack. Also developed classifiers to analyze nodule level malignancy and emphysema. Built POC for identifying cancerous lung nodules from **Radiomics data**. Streamlined prototyping and testing via parallelization of the data pre-processing pipeline (patch extraction and clean-up from CT scans).
- **Google Summer of Code** — Portland State University
  Software Developer Intern — *May 2016 - August 2016*
  - **Cyvlfeat**: Designed and developed 12 new features for a **high-performance Python/Cython wrapper** of computer vision library, VLFeat. (Added algorithms specializing in image understanding and local features extraction and matching such as LBP, SIFT, hierarchical k-means, SLIC). Built **unit and integration tests** using Python's Nose test suite.

Projects

- **Computer Vision**: Built an **attention mechanism** in form of **Region Proposal network** (RPN) for Object detection task. This RPN later served as a backbone for **MaskRCNN** with object detection heads of FasterRCNN and a parallel **mask segmentation** branch. Implemented **vectorized ROIAlign** for FPN-ROI Mapping.
  Developed **YOLO** pipeline (end-to-end) for **object detection**, with a **Non Maximum Suppression** post-processing module to filter most precise detections. Built a semi-automated **optical flow based tracker** for real-time videos.
- **Computational Linguistics**: Using **vector space models**, developed a framework to compare the correlation for human judgments of similarity to the vector similarities. Working on **Bilingual Named Entity Recognition** using Bi-LSTM CRF and Self Attention.
- **Learning Visual control for Car Racing**: Implemented a **Fully connected Deep Q-network** and achieved an average reward of 210.92 for 10 evaluation steps. The best performing model had 70,475 parameters and trained for **only 570 episodes**. Explored other methods like DQNs with dropout and **Proximal Policy Optimization**.
- **Multi-threaded web server and Service framework**: A Java based web **HTTP 1.1 compliant web server** developed with **custom** implementations of underlying Blocking Queue and Thread Pool. Later merged it with a custom-built web service framework which emulates the behaviour of **Java Spark**.
- **Web crawler and XPath Engine**: Developed a **multithreaded web crawler** with a custom XPath Parser and to query and store matched HTML, XML documents into a persistent data store.
- **Search Engine**: Developing Map-Reduced Based Indexer from scratch which will later be integrated with a custom crawler, Page-Rank module and Search Engine UI. (Team size: 4)