Sandipan Choudhuri

[+1] 480-300-9385 | s.choudhuri@asu.edu | s-choudhuri.github.io | s-choudhuri | S-Choudhuri

EDUCATION

PHD | Computer Science — USA | Aug 2017 - Summer 2023 [tentative]
Dissertation: "Domain Adaptation in Unconstrained Label Spaces." — CGPA: 4/4
Arizona State University

MASTERS | Computer Science & Engineering — India | Aug 2013 - Jul 2015
Thesis: "A Quality-Concordance Measure-Based Approach to Edge Detection." — CGPA: 8.44/10
Jadavpur University

BACHELORS | Computer Science & Engineering — India | Aug 2009 - Jul 2013
West Bengal University of Technology — CGPA: 8.60/10

PROFESSIONAL EXPERIENCE

ROVICARE | Machine Learning Intern — USA | May 2022 – Aug 2022

- Developed a layout-agnostic pipeline for seamless extraction of selective information from scanned patient intake forms, using **Azure Form Recognizer Service** and deep-learning routines, with **92%** field-value extraction accuracy.
- Built an efficient word search module for scanned records, with **Azure Cognitive Service**, to expedite the patient intake process.
- Deployed machine learning models as **Python RESTful APIs** on the **Azure App Services** with **Docker** runtime.
- Streamlined **HubSpot** operations by automating deal pipeline stages, developing ticketing system add-ons for analyzing support responses to find actionable insights, and devising an automated contract filling system through **PandaDoc** integration.

INTERBIZ | Machine Learning Intern — USA | May 2021 - Aug 2021

- Reduced the processing time for auto insurance collision claims by **45%** through the development of category detectors and form-processors for the scanned estimates, using **Google Document AI**, **Google Cloud Vision API**, and deep-learning modules.
- Developed a health insurance claim form parser with **Google Cloud Vision API** and field-value pair extractors, yielding **94%** extraction accuracy. Communicated with the stakeholders for requirements gathering, workflow designing, and documentation.
- Built a **Google Cloud Web Application** with **OAuth 2.0 framework** authentication (conforming to HIPAA security and privacy standards) for secured access to healthcare documents from **Google Drive** and **Google Cloud Storage**.

NETXT LAB | Research Assistant — USA | Aug 2019 – Present

- Developing predictive models for Arizona Health Care Cost Containment System (AHCCCS) to combat the state's opioid crisis.
- Built a framework for forecasting future opioid overdose incidence volumes over a geographic region and strategically determining optimal locations (demand points) for setting up Medically Assisted Treatment (MAT) facilities under budgetary constraints.
- Evaluated the effectiveness of Twitter data as a viable indicator of opioid overdose incidence volumes by developing a pipeline for their characterization and analysis (using **RoBERTa**). A framework for detecting illicit online opioid sales is currently underway.

CMATER LAB | Junior Research Fellow — India | Aug 2015 – Mar 2017

- Served as the technical lead for the **Breast Cancer Diagnostic System** development program on **Fine-Needle Aspiration Cytology** images; duties involved research and development, communicating with the expert groups at diagnostic units, and the project management team at Jadavpur University for requirements gathering, data acquisition, and configuring access controls.
- Conducted research, development and evaluation of models for **Object Localization, Semantic Contour Detection** and **Image Segmentation** tasks on natural scene (BSDS500, PASCAL-VOC, Stanford B.) and near-Infrared image (RGB-NIR Scene) datasets.

SKILLS

Research	Domain Adaptation (Closed-Set/Partial/Open-Set/Source-Free) • Contour Detection • Object Localization Semantic Image Segmentation • Unsupervised Learning • Time-Series Analysis
Platform	Google Cloud Platform (GCP) • Microsoft Azure
Language	Python • Java • C++ • C • SQL • JavaScript • HTML • CSS
Technology	Docker • Apache Spark • Tableau • Matlab • Git • ClickUp
Library	Pytorch • Keras • Pandas • Scikit-learn • OpenCV • Apache Spark MLlib • Seaborn • Flask • D3.js

AWARDS & SERVICES

- Recipient of the **Engineering Graduate Fellowship** for *"strong academic work and research progress."*
- Recipient of the **Department Travel Grants**.
- **Top 0.5%** of *0.25 million applicants* who took the *2013 Graduate Aptitude Test in Engineering (GATE-2013)* in Computer Science (conducted by the Indian Ministry of Human Resource Department).
- Served as the **Web chair** for the *International Workshop on Network Science for Quantum Communication Networks, 2022* **(NETSCIQ-COM - INFOCOM)**. Duties involved designing, updating, and stress-testing the conference website, managing submissions, providing tech support, and reviewing workshop live streams.
- **Co-wrote proposals** and was awarded **AHCCCS Research Grant** for the *Arizona State Opioid Response Data Project*.
- **Teaching** and **Research Awards reviewer** for the *Graduate and Professional Student Association* at Arizona State University.

PUBLICATIONS

ASILOMAR 2022	Domain-Invariant Feature Alignment Using Variational Inference For Partial Domain Adaptation **S.Choudhuri**, S. Adeniye, H Venkateswara, A.Sen (accepted)
ADVML@KDD 2022	Coupling Adversarial Learning with Selective Voting Strategy for Distribution Alignment in Partial Domain Adaptation **S.Choudhuri**, H Venkateswara, A. Sen
GLOBECOM 2021	Optimal Cost Network Design for Bounded Delay Data Transfer from PMU to Control Center A. Sen, S. Roy, K. Basu, S. Adeniye, **S. Choudhuri**, A. Pal
DRCN 2020	Structural Dependency Aware Service Chain Mapping for Network Function Virtualization A Sen, **S Choudhuri**, K Basu
ASILOMAR 2020	Partial Domain Adaptation Using Selective Representation Learning For Class-Weight Computation **S.Choudhuri***, R. Paul*, A.Sen, B.Li, H Venkateswara
RDMLDA 2019	Combining Multilevel Contexts of Superpixel using Convolutional Neural Networks to Perform Natural Scene Labeling A Das, S Ghosh, R Sarkhel, **S Choudhuri**, N Das, M Nasipuri
EPIDAMIK@KDD 2018	Identification of At-Risk Groups for Opioid Addiction Through Web Data Analysis K Basu, **S Choudhuri**, A Sen, A Majumdar, D Dey
MMTC 2018	User Satisfaction-Driven Bandwidth Allocation for Image Transmission in a Crowded Environment **S Choudhuri**, K Basu, A Sen
IJPRAI 2018	Object Localization on Natural Scenes: A Survey **S Choudhuri**, N Das, R Sarkhel, M Nasipuri
FICTA 2017	A Quality-Concordance Metric Based Contour Detection by Utilizing Composite-Cue Information and Particle Swarm Optimisation **S Choudhuri**, N Das, M Nasipuri
ICACCI 2016	A Multi-Cue Information Based Approach to Contour Detection by Utilizing Superpixel Segmentation **S Choudhuri**, N Das, S Ghosh, M Nasipuri

NOTABLE PROJECTS

TIME-SERIES ANALYSIS

PYTHON, STATSMODELS, PANDAS, SCIKIT-LEARN, SEABORN

Developed a pipeline for predicting opioid overdose incidence volumes in a future time frame using time-series analysis and forecasting over *Arizona EMS data*, *opioid overdose records* and *demographic* information. Tasks involved time-series decomposition, stationarity testing using *unit-root tests*, transforming the time series to stationary, testing for seasonality, addressing missing values, computing partial autocorrelation and estimating the series forecastability, running *Granger Causality tests* to gauge the effectiveness of the time series in forecasting another, and developing forecasting models using *ARIMA*, *SARIMA*, *SARIMAX*, *VARMA*, and *VARMAX*.

DEEPVID++

PYTHON, KERAS, FLASK, D3, JAVASCRIPT, KNOWLEDGE DISTILLATION

The reliability of a deep-learning model is majorly dependent on the interpretability of its internal operations. Citing this, the *DeepVID* framework is improved to train a basic network using a complex deep-neural model such that its behavior is visually interpretable. This is accomplished by utilizing *variational information* to generate neighbors around data instances and training a linear network to replicate the complex network's behavior on the generated samples through *knowledge distillation*.

DOMAIN ADAPTATION FRAMEWORKS

PYTHON, PYTORCH, KERAS, ADVERSARIAL LEARNING

Developed frameworks for cross-domain image translation, data augmentation and classification tasks under a domain adaptation setup, with *constrained* and *unconstrained* label-space assumptions between the source and target domains (*closed-set*, *partial* and *open-set* scenarios); built adaptation models for scenarios where access to source data is prohibited during the adaptation process (*source-free* scenario).

OBJECT LOCALIZATION

PYTHON, KERAS, OPENCV, MATLAB, COMPUTER VISION

Formulated an approach for object localization and semantic image segmentation in natural scenes, using image *super-pixelization* and *deep supervised classification*. Context-based information is captured through patches formed by the first and second neighbor super-pixels (atomic units of an image formed by a perceptual grouping of pixels). These are then mapped to the corresponding object labels using deep classifier networks. A consensus-based labeling strategy is subsequently employed by ensembling the output probabilities at different scales through multiple voting routines.

IMAGE SEGMENTATION

PYTHON, PYTORCH, BIO-MEDICAL IMAGE PROCESSING

Developed a framework for polyp segmentation in optical colonoscopic video sequences by coupling *deep-siamese networks* with a clustering policy using *balanced entropy-based random walks* to identify informative frames and segment out existing polyps (precursors to colon cancer).

SEMANTIC CONTOUR DETECTION

PYTHON, KERAS, C++, OPENCV, GAME THEORY, COMPUTER VISION

A multi-scale feature-based contour detection approach is presented here from a game-theoretic standpoint. A mixed strategy Nash Equilibrium is searched in a two-player game (*texture* and *color*-based feature extractors acting as players, with strategies involving *extraction at different scales*). The objective centers around finding a weighted balance of these strategies to minimize the false positives (highlighting spurious edges) and false negatives (missing salient contours).

TEACHING

Foundation of Algorithms (TA[1], Spring & Fall 2018-2021) • Data Mining (TA[1], Fall 2019) • Object-Oriented Programming and Data Structures (TA[1], Fall 2017, Spring 2018) • Principles of Programming (TA[1], Fall 2017)

RELEVANT COURSEWORK

Artificial Intelligence • Deep Learning • Statistical Machine Learning • Image Processing • Natural Language Processing • Data Visualization • Pattern Recognition • Vision & Language Frontiers • Data Mining • Information Assurance & Security • Soft Computing • Game Theory • Mobile Computing

[1]TA: Teaching Assistant