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 - 15th Aug 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 — Tempe, USA | May 2022 – Aug 2022

- Leveraged **Azure AI Services** and **image-processing** routines to develop a framework for seamless extraction of requisite entities from scanned patient intake forms with **92%** accuracy, expediting the decision-making process for patient intake by **64%**.
- Engineered a word localizer module using **Azure Cognitive Service** to render scanned patient files searchable for keywords.
- Streamlined **HubSpot** operations for enhancing in-house productivity: built ticketing system add-ons to analyze support requests for actionable insights, automated contract filling through **PandaDoc** integration, and created workflows in deal pipelines.
- Deployed machine learning models as **Python RESTful APIs** on the **Azure App Services** with **Docker** runtime.

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

- Developed a document parser for batch-processing poorly-digitized health insurance forms with **Google Cloud Vision** and **region-proposal networks** with an extraction accuracy of **96%**, translating to an improvement in the turnaround time by **72%**.
- Reduced the processing time for auto insurance collision claims by **49%** by developing estimate identifier and form-processor models for the scanned estimates, using **Google Cloud Services** and **deep-siamese networks**.
- 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, ARIZONA STATE UNIVERSITY | Research Assistant — Tempe, USA | Aug 2019 – Present

- Addressed the data annotation overhead by proposing **transfer-learning frameworks** with realistic label-set overlap assumptions; achieved the highest accuracies in **67%** of the adaption challenges against state-of-the-art models on benchmark datasets.
- Designed a time-series pipeline for jointly estimating the **geospatial distribution of future opioid overdose incidences** and **optimal sites** to establish MAT [1] facilities under budgetary constraints; an AHCCCS[2] initiative to **combat the state's opioid crisis**.
- Evaluated the effectiveness of Twitter data to monitor the opioid crisis through characterization using **RoBERTa** language model.

CMATER LAB, JADAVPUR UNIVERSITY | Junior Research Fellow — Kolkata, India | Aug 2015 – Mar 2017

- Technical lead in a production-level **Breast Cancer Diagnostic System** development program on Fine-Needle Aspiration Cytology images. Duties involved conducting research and evaluation, mentoring the development team, communicating with the expert groups at diagnostic units for requirements gathering, data acquisition, and configuring access controls.
- Conducted exploratory analysis on challenges of **Scene Labelling, Contour Detection**, and **Image Segmentation**. Addressed them by proposing novel machine-learning frameworks for RGB and near-infrared image datasets and publishing findings.

COURSEWORK & SKILLS

Course[3]	Deep Learning • Machine Learning • Statistical Learning • Computer Vision • Data Mining • Data Visualization
Teaching[3]	Data Mining • Foundation of Algorithms • Object-Oriented Programming & Data Structures
Research	Domain adaptation (Unsupervised • Closed-set • Partial • Open-set • Source-Free) • Natural Scene Labelling
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 • NumPy • Apache Spark MLlib • Seaborn • Flask • D3.js

[1] MAT: Medically Assisted Treatment [2] AHCCCS: Arizona Health Care Cost Containment System [3] Selected graduate AI courses/Teaching Assistantships

COURSE PROJECTS

DOMAIN ADAPTATION IN UNCONSTRAINED LABEL-SPACE PYTHON, PYTORCH, SCIKIT-LEARN

Proposed a framework for unsupervised target classification under partial adaptation → Conditional distribution alignment achieved by coupling *domain-adversarial neural network* and an iterative category-importance estimator with *intra* and *inter* class distance optimization → Bagged highest accuracies across benchmark models on *Office-31*, *Office-home*, *VISDA2017*, and *ImageNet-Caltech*.

DNN VISUALIZATION PYTHON, PYTORCH, SCIKIT-LEARN

Designed a visualization framework to interpret the functioning of a deep-neural network → Utilized *variational autoencoder* to generate neighbors around data instances and trained a *linear network* to probe and replicate the *complex network's* behavior on the generated samples using *knowledge distillation* → Examined platform's efficacy using *ResNet* and *VGG* models on *MNIST* and *SVHN* datasets.

TIME-SERIES FORECASTING PYTHON, STATSMODELS, PANDAS, SCIKIT-LEARN, SEABORN, TABLEAU

Developed a pipeline for estimating geospatial distribution of future opioid incidences using time-series analysis on *Arizona EMS* and *demographic* data → Tasks involved time-series decomposition, stationarity and seasonality testing, estimating series forecastability through running *Granger causality tests*, and developing forecasting models using *ARIMA*, *SARIMA*, *RNN* and *uni/bi-directional LSTM*.

POLYP SEGMENTATION PYTHON, PYTORCH, OPEN-CV, BIO-MEDICAL IMAGE PROCESSING

Proposed a method to bypass the necessity for constant attention during optical colonoscopy monitoring → Polyp regions estimated from *informative frames* (*informativeness* determined with supervision) using an *iterative entropy-based clustering policy* and a *deep-siamese network* trained on contrastive loss → A *3%* improvement in *mIOU*[1] witnessed over the benchmark *UNet* on a real-world dataset.

NATURAL SCENE LABELLING PYTHON, KERAS, OPENCV, MATLAB, COMPUTER VISION

Outlined a multi-scale solution to capturing context information for scene labelling → *Super-pixel* groups at different neighborhood scales mapped to corresponding object labels using deep classifiers. A *consensus-labeling strategy* employed by ensembling the output probabilities through multiple voting routines → Outperformed *state-of-the-art super-pixel-based methods* on *Stanford B.* dataset.

AWARDS & SERVICES

- Recipient of **Engineering Graduate Fellowships** at *Arizona State University* for *"strong academic work and research progress."*
- Recipient of **Ph.D. Conference Fellowships** from the *School of Computing and Augmented Intelligence*, *Arizona State University*.
- **Top 0.5%** of *0.25 million applicants* who took the National *Graduate Aptitude Test in Engineering-2013* (India) in Computer Science.
- Served as the **Web chair** for the *International Workshop on Network Science for Quantum Communication Networks*, *2022* **(NETSCIQ-COM - INFOCOM)**. Duties involved designing conference website, reviewing submissions and workshop live streams.
- **Co-wrote research grants** and was awarded **AHCCCS Grant** for the *Arizona State Opioid Response Data Project*.
- **Teaching** and **Research Awards reviewer** for the *Graduate & Professional Student Association* at *Arizona State University*.
- **Reviewer** for IEEE journals and conferences.

PUBLICATIONS

ACSSC 2022 — Domain-Invariant Feature Alignment Using Variational Inference For Partial Domain Adaptation
S.Choudhuri, S. Adeniye, H Venkateswara, A.Sen

AIA 2022 — Distribution Alignment Using Complement Entropy Objective and Adaptive Consensus-Based Label Refinement For Partial Domain Adaptation
S.Choudhuri, S. Adeniye, A.Sen

ADVML@KDD JCCE 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

ACSSC 2020 — Partial Domain Adaptation Using Selective Representation Learning For Class-Weight Computation
S.Choudhuri*, R. Paul*, A.Sen, B.Li, H Venkateswara

[1] mIOU: mean Intersection Over Union

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