

# Salam Michael Singh

Email : salammichaelcse@gmail.com

Linkedin: <https://www.linkedin.com/in/michael-salam-93b52074/>

Mobile : +91-7005568277

Github: <https://github.com/masonreznov/>

Blog: <https://masonreznov.github.io/>

## EDUCATION

---

- **National Institute of Technology Silchar** Silchar, Assam  
*PhD, Computer Science and Engineering* Aug 2019 - Current (Expected Submission: November 2022)
- **National Institute of Technology Silchar** Silchar, Assam  
*M.Tech, Computer Science and Engineering; 8.05/10 CGPA* Aug 2017 - May 2019
- **National Institute of Technology Arunachal Pradesh** Yupia, Arunachal Pradesh  
*B.Tech, Computer Science and Engineering; 7.39/10 CGPA* Aug 2012 - May 2016

## TECHNICAL SKILLS

---

- **Languages:** Python, C, Java, BASH Scripting, MATLAB
- **ML/DL:** Pytorch, Numpy, Trax, Scikit-Learn, NLTK, Gensim, Huggingface, Moses, spacY, Lightning
- **Data Science:** Pandas, Matplotlib, Plotly, Jupyter, MS Excel
- **Others:** L<sup>A</sup>T<sub>E</sub>X, MS Word, Git, HTML/CSS/JS, MySQL, Streamlit, Flask, Gradio, Scrapy, BeautifulSoup

## RESEARCH INTERESTS

---

Machine Translation, Natural Language Processing, Computer Vision, Deep Learning

## WORK EXPERIENCE

---

- **National Institute of Technology Silchar** Silchar, Assam  
*Teaching Assistant* August 2020 - Current
  - Natural Language Processing — Introduction to Programming — Object-Oriented Programming
- **National Institute of Technology Silchar** Silchar, Assam  
*Graduate Teaching Assistant* August 2018 - June 2019
  - Introduction to Programming — Data Structures

## RESEARCH EXPERIENCE

---

- **Low Resource Machine Translation** NIT Silchar, Assam  
*Advisor: Dr. Thoudam Doren Singh* August 2020 - Current
  - **Multilingual Machine Translation:**
    - Proposed a multilingual cross-lingual word embedding as the initialiser for a multilingual neural machine translation system. Improved over the supervised as well as the vanilla multilingual model for machine translation of the low resource languages.
    - Explored the effectiveness of transfer learning from large scale pretrained multilingual models such as **mBART** and **mT5** for finetuning on low resource language pairs absent during the pretraining stage.
  - **Unsupervised Machine Translation:**
    - Implemented a transformer based unsupervised neural machine translation with shared encoder and language specific decoders for low resource English and Manipuri translation. Improved over the RNN based system by +3 BLEU score and the cross-lingual as a prior is found to be effective as against the pretrained language model as prior for this language pair.
    - Participated in WMT 2020 shared task for unsupervised machine translation where MASS pretraining and finetuning objective was used along with a noisy supervised training on the synthetic data.
  - **Multimodal Machine Translation:**
    - Proposed a multi-modal NMT in a multilingual setup.
    - Used the multiple captions as a paraphrase and jointly trained which addressed the lexical diversity.
- **Multi-Modal Word Sense Disambiguation** NIT Silchar, Assam  
*Advisor: Dr. Samir Kr. Borgohain* August 2018 - June 2019
  - Proposed and built a machine learning based multi-modal (text and image) word sense disambiguation model.

## PROJECTS

---

- **Unsupervised Neural Machine Translation Tool:** Developed a PyTorch based unsupervised neural machine translation tools with the options of transformer or RNN (LSTM and GRU) based encoder-decoder.
- **TRAX-NER:** Developed a TRAX based named entity recognition tool based on a sequence-to-sequence model.
- **Parallel Data Crawler:** Developed a Scrapy based web crawler to extract parallel data for English and Manipuri translation.

## PROFESSIONAL SERVICES

---

- **Programme Committee:**
  - Conferences/workshops: MMTLRL 2021 @ RANLP 2021
- **Volunteering:**
  - Conferences: ICON 2021, BigDML 2019
  - Workshops: MMTLRL 2021 @ RANLP 2021
- **Reviewing:**
  - Journals: Multimedia Tools and Applications, Journal of Electronic Imaging (JEI)
  - Conferences/workshops: ICON 2021, MMTLRL 2021 @ RANLP 2021
- **Membership:**
  - ACL

## LIST OF PUBLICATIONS

---

- **Journals**
  1. **Salam Michael Singh** and Thoudam Doren Singh. 2022. Low resource machine translation of English-Manipuri: A semi-supervised approach. Expert Systems with Applications. (**SCIE**, IF=**8.665**)
  2. **Salam Michael Singh** and Thoudam Doren Singh. 2022. An empirical study of low-resource neural machine translation of Manipuri in multilingual settings. Neural Computing and Applications. (**SCIE**, IF=**5.130**)
  3. Arpita Dutta, **Salam Michael Singh** and Samir Kumar Borgohain. 2021. Removal of Ambiguity of Noun Using Multimodal Approach. Indian Journal of Computer Science and Engineering (IJCE). (SCOPUS)
- **Conferences/workshops/shared tasks:**
  1. **Salam Michael Singh**, Loitongbam Sanayai Meetei and Thoudam Doren Singh. 2021. Multiple Captions Embellished Multilingual Multi-Modal Neural Machine Translation. Proceedings of the First Workshop on Multimodal Machine Translation for Low Resource Languages MMTLRL 2021 at **RANLP 2021**.
  2. **Salam Michael Singh** and Thoudam Doren Singh. 2020. Unsupervised Neural Machine Translation for English and Manipuri. In the third Workshop on Technologies for MT of Low Resource Languages (LoResMT 2020) at **AAACL-IJCNLP 2020**.
  3. **Salam Michael Singh**, Thoudam Doren Singh, and Sivaji Bandyopadhyay. 2020. The NITS-CNLP system for the unsupervised machine translation task at WMT 2020. In the Fifth Conference on Machine Translation at **EMNLP 2020**.
  4. **Salam Michael Singh**, Loitongbam Sanayai Meetei, Alok Singh, Thoudam Doren Singh and Sivaji Bandyopadhyay. 2021. On the Transferability of Massively Multilingual Pretrained Models in the Pretext of the Indo-Aryan and Tibeto-Burman Languages. In Proceedings of the 18<sup>th</sup> International Conference on Natural Language Processing (**ICON 2021**).
  5. Loitongbam Sanayai Meetei, **Salam Michael Singh**, Alok Singh, Thoudam Doren Singh and Sivaji Bandyopadhyay. 2021. An Experiment on Speech-to-Text Translation Systems for Manipuri to English on Low Resource Setting. In Proceedings of the 18<sup>th</sup> International Conference on Natural Language Processing (**ICON 2021**).
  6. Alok Singh, Loitongbam Sanayai Meetei, **Salam Michael Singh**, Thoudam Doren Singh and Sivaji Bandyopadhyay. 2021. An efficient keyframes selection-based framework for video captioning. In Proceedings of the 18<sup>th</sup> International Conference on Natural Language Processing (**ICON 2021**).