

Shreekantha Nadig

shreekantha.nadig@iiitb.ac.in

<https://github.com/sknadig>

| <https://vak.ai/>

| <https://www.linkedin.com/in/sknadig/>

EXPERIENCE

SENIOR SPEECH RECOGNITION ENGINEER | DIALPAD

Toronto | Bengaluru | Dec 2019 – Present

- Lead the team in transitioning the product from Kaldi to next-gen ASR
- Architected a next-gen Speech Recognition product end-to-end from R&D to production
 - Large scale: Scaled the product to transcribe 1M+ hours of audio per day
 - Toolkit agnostic: Built an ASR engine that can accept AM and LM from different ASR OSS toolkits
 - Better performance: Optimized the network architecture and decoder to have better WER, F1 scores
 - Lower latency: Implemented a decoder that has lower runtime (RTF) and latency (UX) than Kaldi
 - More efficient: Optimized the model runtime to have better throughput and reduce compute cost
 - Multilingual: Collaborated with the team to scale to multiple languages and handle code-switching
 - Team building: Up-skilling the team to use the latest toolkits and develop next-gen ASR models
- Developing further improvements to next-gen ASR Engine with K2/icefall and Nvidia NeMo
- Developed a LM customization pipeline to improve ASR performance on customer-specific datasets
- Developed low latency redis integration for synchronization of transcripts from different call legs
- Built and benchmarked various end-to-end ASR architectures with CTC, Attention-based Encoder-Decoder (AED), Transducer, Transformer and Conformer models with hybrid ASR models and external ASR services
- Developed interfaces for shallow fusion of multi-level (sub-word and word) RNNLMs and n-gram LMs
- Developed methods to bias the models towards a list of keywords, resulting in an absolute WERR of 7%
- Automated the data preparation pipeline for training ASR models, reducing the turnaround time for experiments and increasing productivity of the team
- Developed pronunciation-assisted sub-word models using fast-align, GIZA++ and Pynini, resulting in an absolute WERR of 3% compared to BPE sub-words
- Post training quantization of ASR models to achieve 50% faster RTF and 75% smaller models on disk
- Implemented the ASR inference in ONNX runtime, reducing the latency by 3x
- Developed performance monitoring techniques for end-to-end ASR models based on RNN-AED and CTC confidence scores, and their use in semi-supervised learning techniques
- Developed better endpoint detection for hybrid models and achieved 4% relative WERR
- Developed a web-app for internal users to query production calls and visualize hypotheses using wavesurfer-js

MACHINE LEARNING INTERN - ASR | OBSERVE.AI

Bengaluru, IN | May 2019 – Aug 2019

- Developed a feature extraction pipeline using tf.signal and tf.data
- Implemented different keyword-spotting (KWS) papers - Deep-KWS, CTC KWS
- Developed methods to convert a custom PyTorch model to TensorFlow
- Deployed the KWS model using TensorFlow serving with an RTF of 0.05 on GPU

SVT ENGINEER | SONUS NETWORKS

Bengaluru, IN | Aug 2015 – Jan 2017

- Worked as a part of Sustaining SVT on Real-Time communication products Sonus Insight (EMS) and SBC
- Developed automated test frameworks in Python, Perl, Linux and Java to test Sonus's EMS product
- Worked with CentOS, RHEL and Solaris to develop and test the EMS and SBC products
- Developed automation tools that reduced testing time from many hours to a couple of minutes

SKILLS

PROGRAMMING Python, Bash, JavaScript, C++

LIBRARIES PyTorch, TensorFlow, Keras, scikit-learn, Django, OpenCV

TOOLKITS NeMo, K2, Kaldi, ESPnet, WeNet

EDUCATION

MS BY RESEARCH - DATA SCIENCE | IIIT - BANGALORE

Bengaluru, IN

Thesis: Multi-task learning in end-to-end attention-based automatic speech recognition

B.E. - TELECOMMUNICATION ENGINEERING | JNNCE (VTU)

Shivamogga, IN

MS THESIS

MULTI-TASK LEARNING IN END-TO-END ATTENTION-BASED AUTOMATIC SPEECH RECOGNITION | ESPNET, KALDI, PYTORCH, TENSORFLOW

Bengaluru, IN

- Developing state-of-the-art systems for end-to-end ASR using Joint CTC and Attention-based models
- Study how pure data-driven models can be blended with knowledge-based models for reducing model complexity, faster training/inference and extracting deeper insights into speech recognition
- Use of ASR toolkits like Kaldi, ESPnet with PyTorch and TensorFlow to build end-to-end ASR models
- Study of various Attention mechanisms and how they can be modelled efficiently for interpretability, explainability of end-to-end models
- Multi-target hybrid CTC/Attention network for joint phoneme-grapheme recognition
- Jointly learning to align and transcribe using attention-based alignment and uncertainty-to-weight losses
- Keyword-spotting using attention-based end-to-end ASR models

PUBLICATIONS

- Riqiang Wang*, **Shreekantha Nadig***, Daniil Kulko, Simon Vandieken, Chia-tien Chang, Seyyed Saeed Sarfjoo, and Jonas Robertson. 2024. "Double Decoder: Improving latency for Streaming End-to-end ASR Models." In Proceedings of the 7th International Conference on Natural Language and Speech Processing (ICNLSP 2024), pages 83–91, Trento. Association for Computational Linguistics.
- Vasundhara Gautam, Wang Yau Li, Zafarullah Mahmood, Frederic Mailhot, **Shreekantha Nadig**, Riqiang Wang and Nathan Zhang, "Avengers, Ensemble! Benefits of ensembling in grapheme-to-phoneme prediction" In Proceedings of the 18th SIGMORPHON Workshop on Computational Research in Phonetics, Phonology, and Morphology, 2021. **(Best overall submission)**
- **Shreekantha Nadig**, V. Ramasubramanian, Sachit Rao, "Multi-target hybrid CTC-Attentional Decoder for joint phoneme-grapheme recognition," 2020 International Conference on Signal Processing and Communications (SPCOM), Bangalore, India, 2020
- **Shreekantha Nadig**, Sumit Chakraborty, Anuj Shah, Chaitanay Sharma, V. Ramasubramanian, Sachit Rao, "Jointly learning to align and transcribe using attention-based alignment and uncertainty-to-weight losses," 2020 International Conference on Signal Processing and Communications (SPCOM), Bangalore, India, 2020. **(Best Student Paper Award – Honorable Mention)**
- Abhijith Madan, Ayush Khopkar, **Shreekantha Nadig**, K. M. Srinivasa Raghavan, V. Ramasubramanian, "Semi-supervised learning for acoustic model retraining: Handling speech data with noisy transcript," 2020 International Conference on Signal Processing and Communications (SPCOM), Bangalore, India, 2020

PRE-PRINT

- Wang Yau Li, **Shreekantha Nadig**, Karol Chang, Zafarullah Mahmood, Riqiang Wang, Simon Vandieken, Jonas Robertson, Fred Mailhot. 2023. "N-gram Boosting: Improving Contextual Biasing with Normalized N-gram Targets." arXiv preprint arXiv:2308.02092

AWARDS

THIRD PRIZE MUCS 2021: MULTILINGUAL AND CODE-SWITCHING ASR CHALLENGES FOR LOW RESOURCE INDIAN LANGUAGES

Bengaluru, IN | Aug 2021

https://github.com/dialpad/mucs_2021_dialpad

Team contributions to multilingual and low-resource ASR for Indian Languages. Benchmarking and open-sourcing various E2E methods and studying effects of channel distortions on language identification.

BEST STUDENT PAPER AWARD – HONORABLE MENTION SPCOM 2020

Bengaluru, IN | Jul 2020

Jointly learning to align and transcribe using attention-based alignment and uncertainty-to-weight losses