Building Foundational Models & Generalizable AI in 6G

ECE Seminar Series

Royal Military College of Canada & IEEE Kingston Section

Hatem Abou-Zeid, PhD
Electrical & Software Engineering Department
University of Calgary, Canada

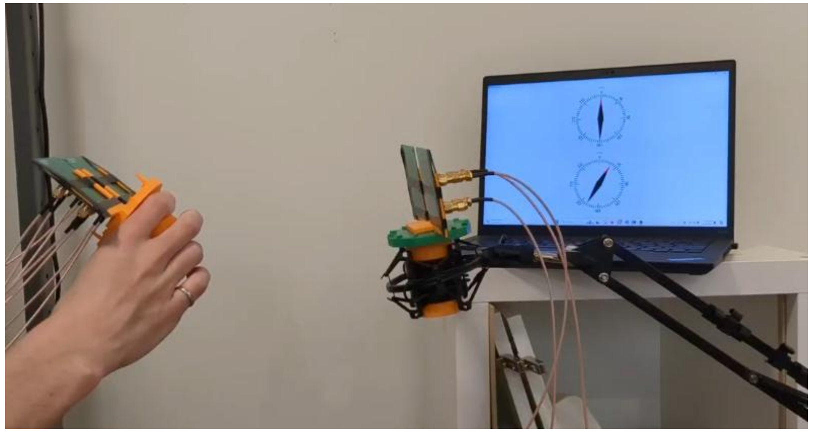
February 2025

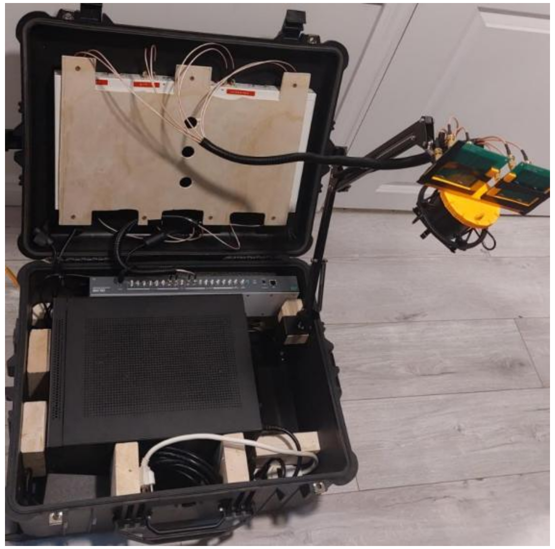
UNIVERSITY OF CALGARY

AI Generalization Challenges in Wireless

Models do not generalize/adapt

When the Software Defined Radio (SDR) gets hot the Angle of Arrival model accuracy went down!





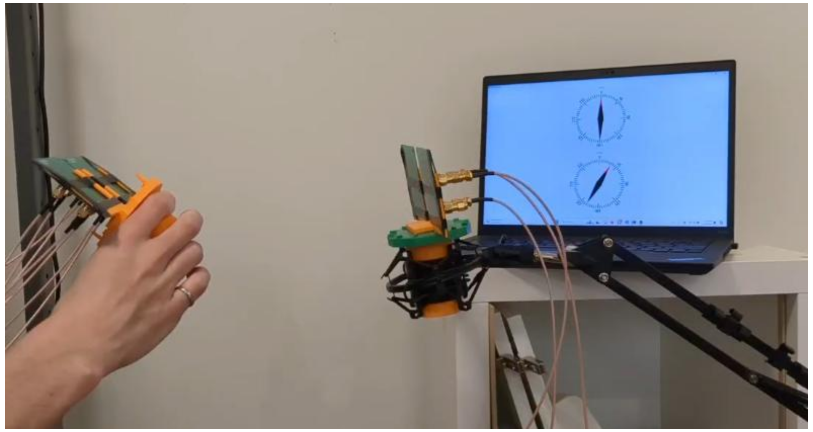
AI Generalization Challenges in Wireless

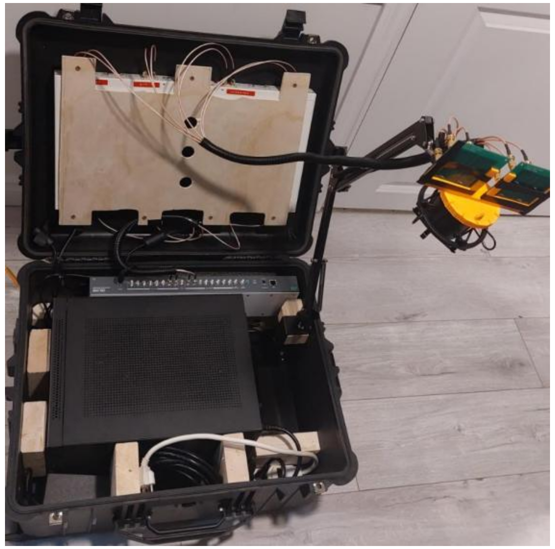
UNIVERSITY OF CALGARY

Models do not generalize/adapt

When the Software Defined Radio (SDR) gets hot the Angle of Arrival model accuracy went down!

Models need lots of data that is labeled





Qoherent

SCHULICH School of Engineering

Why Foundation Models in Wireless?

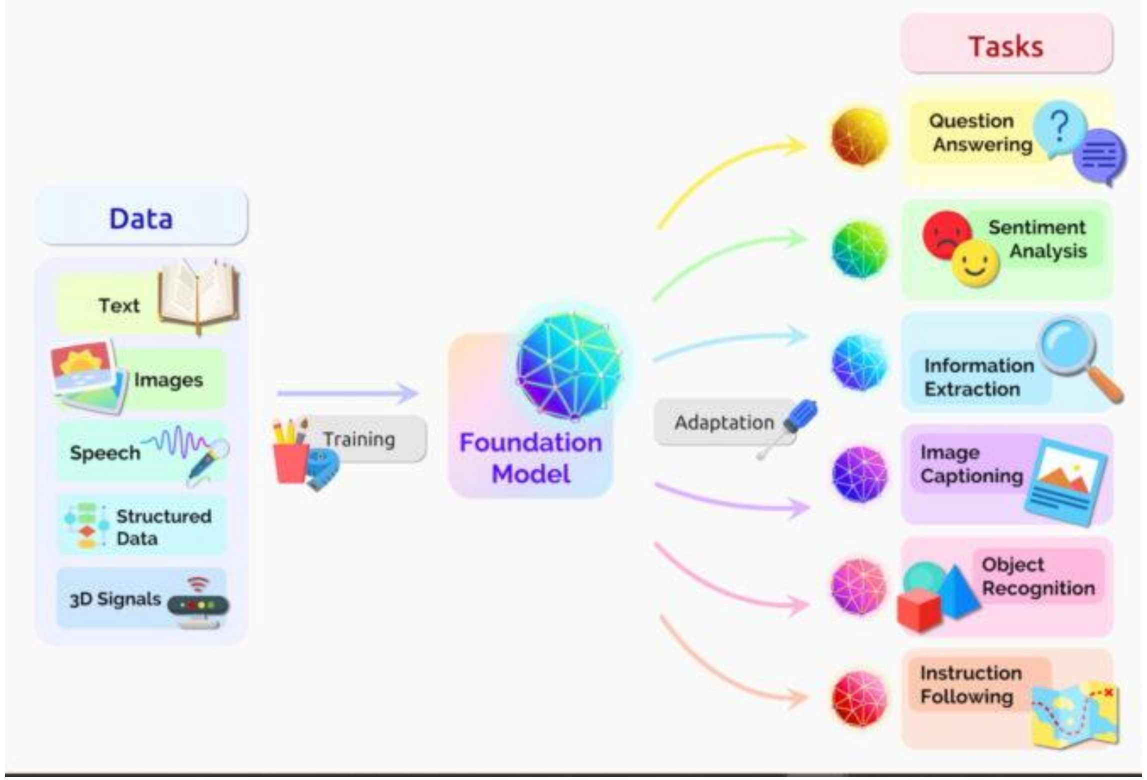
Models do not generalize/adapt

Avoid the need for so many very specific models per task

Models need lots of data that is labeled

Foundation Models (FM)

- Foundation models are a large, pre-trained machine learning models that serve as the basis/foundation for a wide range of downstream tasks.
- ChatGPT is built on a foundation model

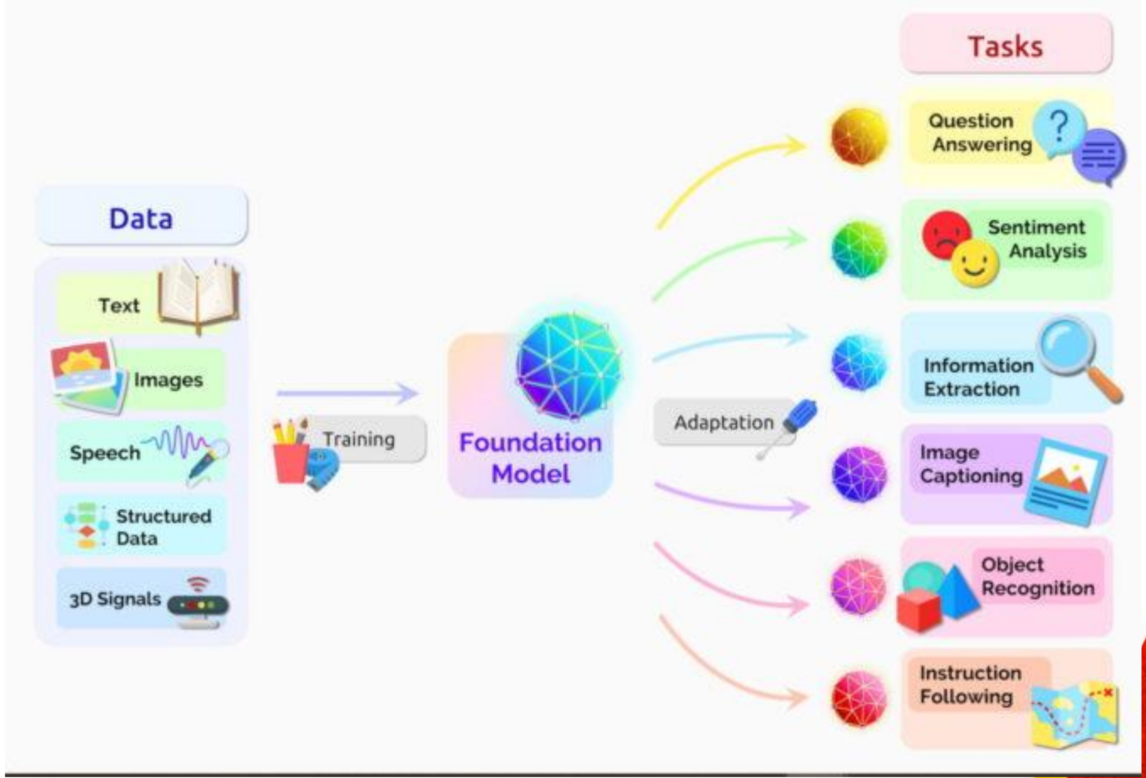


https://blogs.nvidia.com/blog/what-are-foundation-models/

Foundation Models (FM) – Two key properties

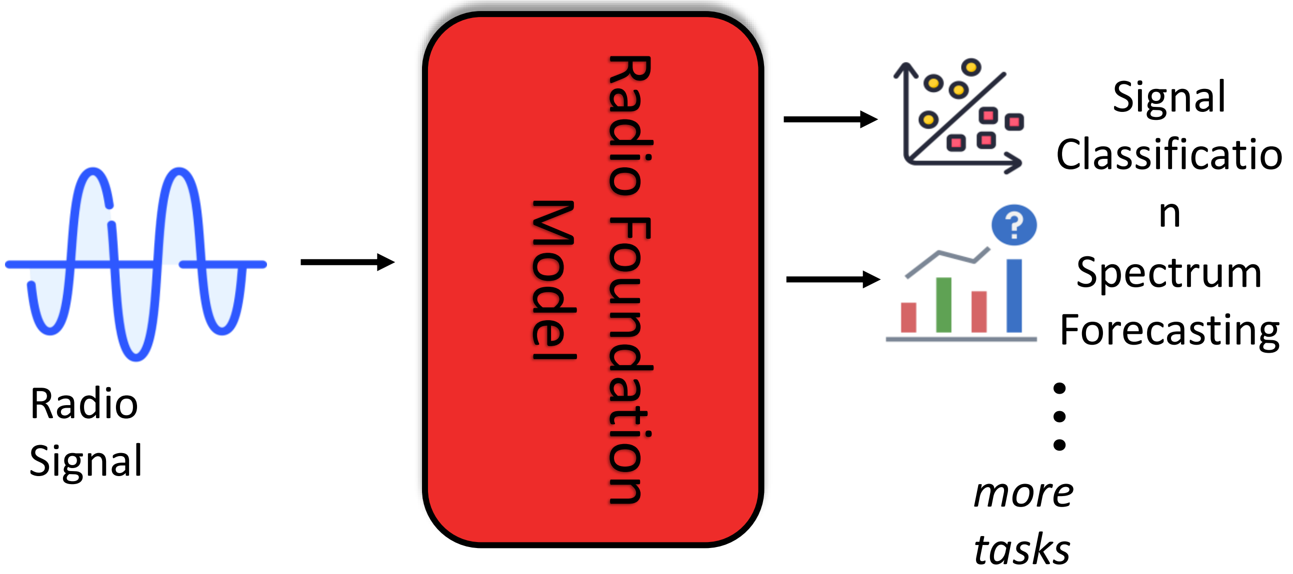
1. <u>Generalization across Tasks</u>:
   - Fine-tuning enables diverse "downstream" tasks with relatively little additional training.

2. Do not require large <u>labeled data for pre-training</u>
   - Use methods such as self-supervised learning



https://blogs.nvidia.com/blog/what-are-foundation-models/

(Why) Build Radio Foundation Models



- **Generalization**: The model has learned from a **diverse range of data**, so it performs well on **new, unseen situations**. It should be able to generalize to different SNRs, bandwidth, frequency ranges, FDD/TDD etc
- **Data Efficiency**: Unlike traditional models that need lots of labeled data, **foundation models perform well with small datasets**.
- **Energy Efficiency**: Instead of training a new model for each task, **one powerful model** can handle multiple tasks.

Talk Overview

1. Generalizable AI for Beam Prediction that Adapts to RF Front Ends
   - A solution using Prototypical Networks

2. Building a Radio Foundation Model using Transformer Architectures
   - Self-supervised pre-training method
   - Evaluation on three downstream tasks (human sensing, localization, RF signal identification)

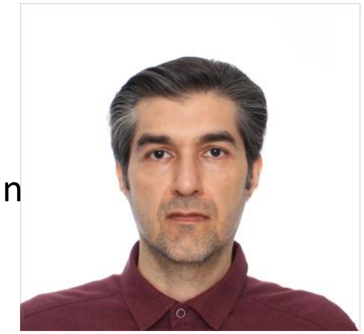
- Research Challenges & Future Research
- Q&A

WAVES Research Group

This research would not have been possible without this amazing group of researchers

Mohammadreza Behboodi Elsayed Mohamed Ogechukwu Kanu Brian Irvine Omar Mashaal Fazal Khan

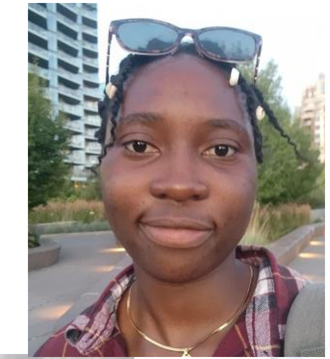
Mohamed Hallaq Sampreet Vaidya Ahmed Aboulfotouh Morvarid Lelanoor



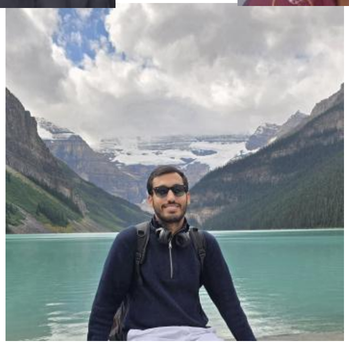
Dr. Bakhshi
PDF & Lab Manager

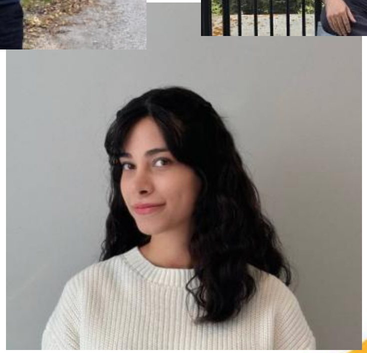
Dr. Ahmad Nagib
VPR Postdoctoral Fellow









https://www.hatem-abouzeid.com/waves-lab

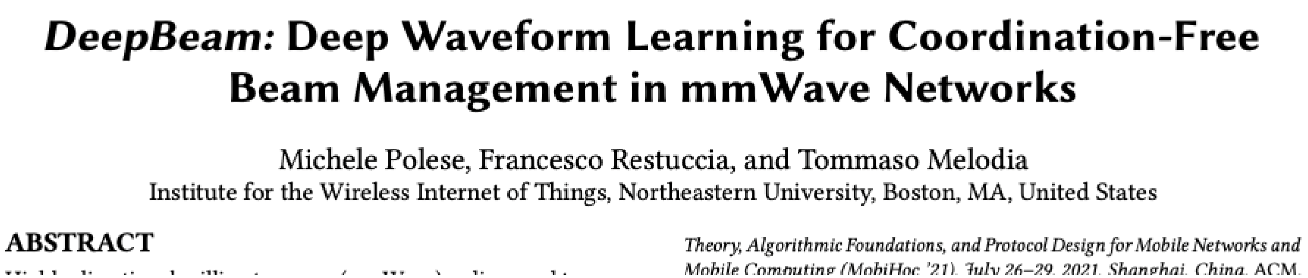
Generalizable AI using Radio Embeddings

Omar Mashaal, PhD Candidate
& Alberta Innovates Fellow

ALBERTA INNOVATES
Qoherent
mitacs
UNIVERSITY OF CALGARY

DeepBeam Overview

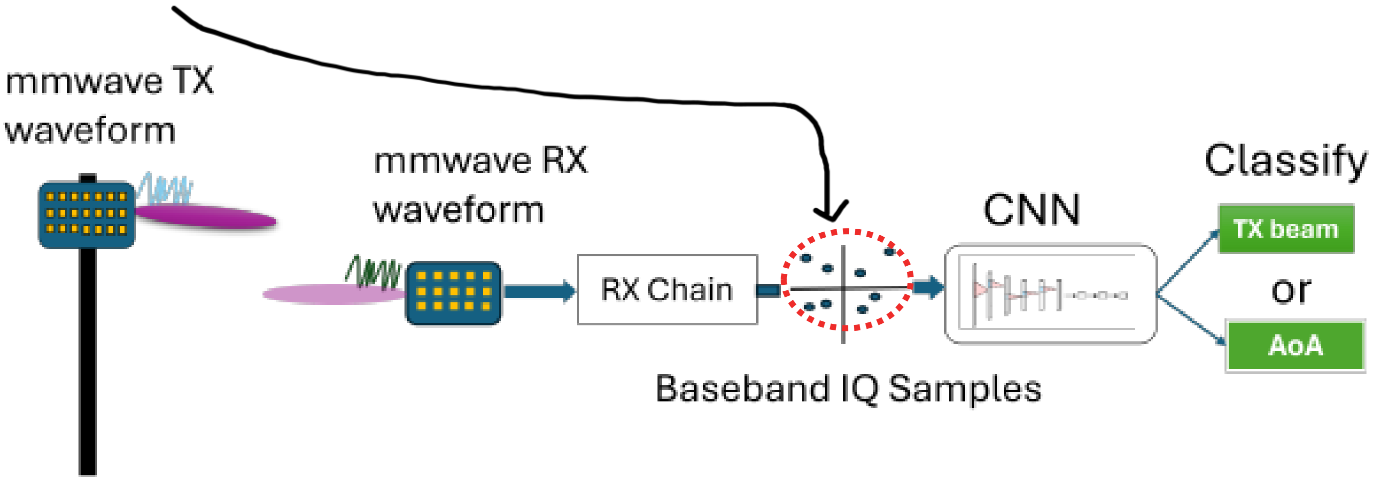
- A framework for waveform-level downlink beam management.
- Infers transmit beams (TXB) without TX-RX coordination. Identifies which of the 24 beams the TX is using
- Processes I/Q samples directly


DeepBeam: Deep Waveform Learning for Coordination-Free Beam Management in mmWave Networks

Michele Polese, Francesco Restuccia, and Tommaso Melodia
Institute for the Wireless Internet of Things, Northeastern University, Boston, MA, United States

ABSTRACT

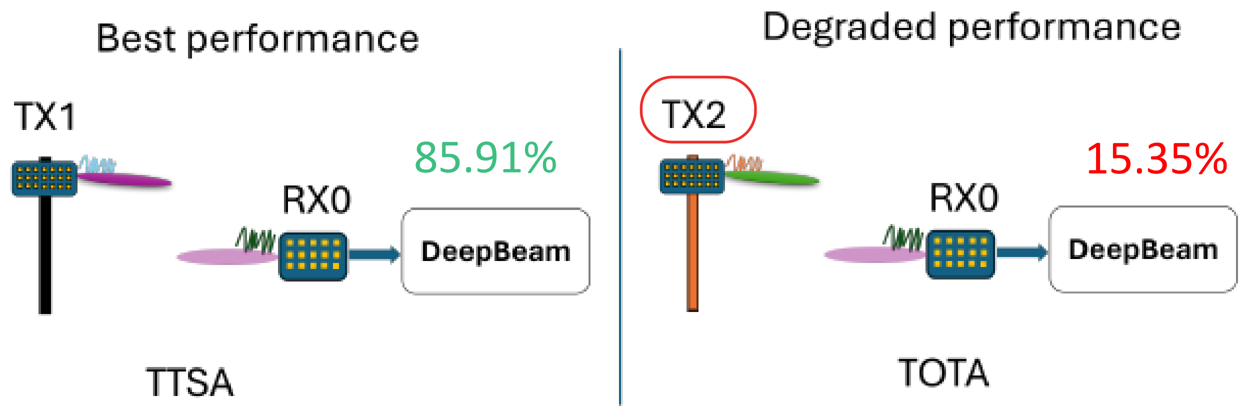
Theory, Algorithmic Foundations, and Protocol Design for Mobile Networks and



M. Polese, F. Restuccia, and T. Melodia. "DeepBeam: Deep waveform learning for coordination-free beam management in mmWave networks." *Proc. Twenty-second International Symposium on Theory, Algorithmic Foundations, and Protocol Design for Mobile Networks and Mobile Computing*. 2021.

The Generalization Challenge: *Different RF Front-ends*

- 4 different SiBeam 60 GHz frontends were used to evaluate generalization
- The accuracy dropped from 85.91% to 15.35% when the the model is deployed with a different RF front end RF front end than the one it was trained on



TTSA: Train and Test on the *Same Antenna (TX1)*
TOTA: Train on One and Test on *Another Antenna (TX2)*

M. Polese, F. Restuccia, and T. Melodia. "DeepBeam: Deep waveform learning for coordination-free beam management in mmWave networks." *Proc. Twenty-second International Symposium on Theory, Algorithmic Foundations, and Protocol Design for Mobile Networks and Mobile Computing*. 2021.

What is non-Generalizable AI?

- The AI model is too *<u>specific to the data</u> it was trained* on
  - Does not generalize to out-of-distribution (OOD) data for the same task or to other related tasks

O. Mashaal, "Towards AI-Native 6G Systems: Robustness and Efficiency" PhD Proposal, UCalgary, 2024.

What is non-Generalizable AI?

Many Causes of Domain Shifts in Wireless

Table 1.1: Illustrative Examples of Causes of Domain Shift in Wireless Signals and Their Direct Impacts

Cause of Domain Shift	Direct Effect on Wireless Signals
Hardware Variations	Power level changes, IQ imbalance, and oscillator frequency drift.
Environmental Changes	Variations in signal propagation paths, shadowing, and multi-path effects.
User Mobility	Doppler shifts, variability in multi-path profiles, and temporal fading.
Interference	Increased noise floor, signal distortion, and potential overlap of frequency bands.

Can we develop wireless AI models that 1) generalize across distribution shifts and 2) rapidly adapt to unseen scenarios with limited labeled data?

O. Mashaal, "Towards AI-Native 6G Systems: Robustness and Efficiency" PhD Proposal, UCalgary, 2024.

ProtoBeam: *Generalization to RF Front-ends*

ProtoBeam: Generalizing Deep Beam Prediction to Unseen Antennas using Prototypical Networks

Omar Mashaal[‡], Elsayed Mohammed[‡], Alec Digby*, Lorne Swersky*, Ashkan Eshaghbeigi*, Hatem Abou-Zeid[‡]
[‡]Department of Electrical and Software Engineering, University of Calgary, Canada
*Qoherent Inc., Toronto, Ontario, Canada

***Abstract*—Deep learning (DL) techniques have recently emerged to efficiently manage mmWave beam transmissions without requiring time consuming beam sweeping strategies. A fundamental challenge in these methods is their dependency on hardware-specific training data and their limited ability to generalize. Large drops in performance are reported in literature when DL models trained in one antenna environment are applied in another. This paper proposes the application of Prototypical Networks to address this challenge – and utilizes the DeepBeam real-world dataset [1] to validate the developed solutions. Prototypical Networks (PN) excel in extracting features to establish class-specific prototypes during the training, resulting in precise embeddings that encapsulate the defining features of the data. We demonstrate the effectiveness of PN to enable generalization of deep beam predictors across unseen antennas. Our approach, which integrates data normalization and prototype normalization with the PN, achieves an average beam classification accuracy of 74.11% when trained and tested on different antenna datasets. This is an improvement of 398% compared to baseline performances reported in literature that do not account for such domain shifts. To the best of our knowledge, this work represents the first demonstration of the value of Prototypical Networks for domain adaptation in wireless networks, providing a foundation for future research in this area.**

Beam Management, Domain Adaptation, Prototypical network, angle-of-arrival, mm-wave.

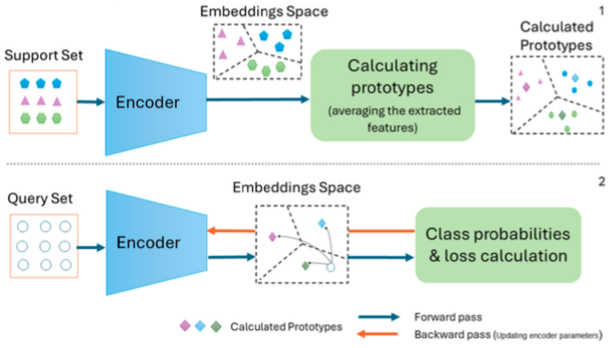
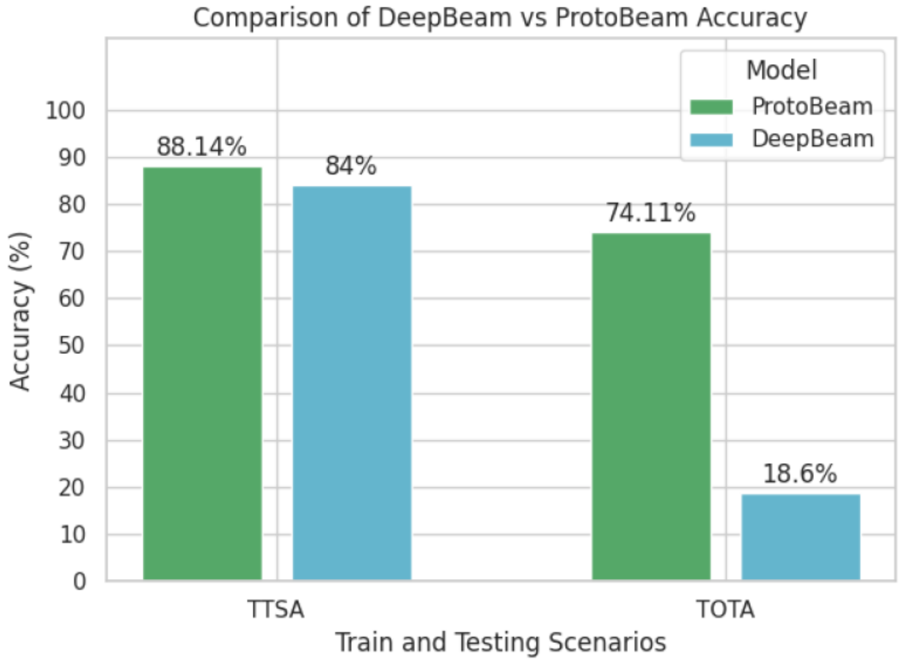


Fig. 1: Prototypical Network Architecture and Training [4].

transmission. Deep learning (DL) has emerged as a powerful tool for refining beam management strategies, enabling dynamic beam prediction and alignment [1]–[3]. DeepBeam [1] leverages deep learning to optimize beam selection using I/Q data. This framework is designed to infer the Angle of Arrival (AoA) and identify the beam used by the trans-



O. Mashaal, et. al. "ProtoBeam: Generalizing Deep Beam Prediction to Unseen Antennas using Prototypical Networks." IEEE GLOBECOM 2024

Prototypical Networks & Few-shot Learning

Prototypical networks for few-shot learning
J Snell, K Swersky, R Zemel - Advances in neural ..., 2017 - proceedi
... We propose **Prototypical Networks** for the problem of **few-shot** c
set, given only a small number of examples of each new class. **Protot**

Prototypical Networks for Few-shot Learning

Jake Snell
University of Toronto*
Vector Institute

Kevin Swersky
Twitter

Richard Zemel
University of Toronto
Vector Institute
Canadian Institute for Advanced Research

Abstract

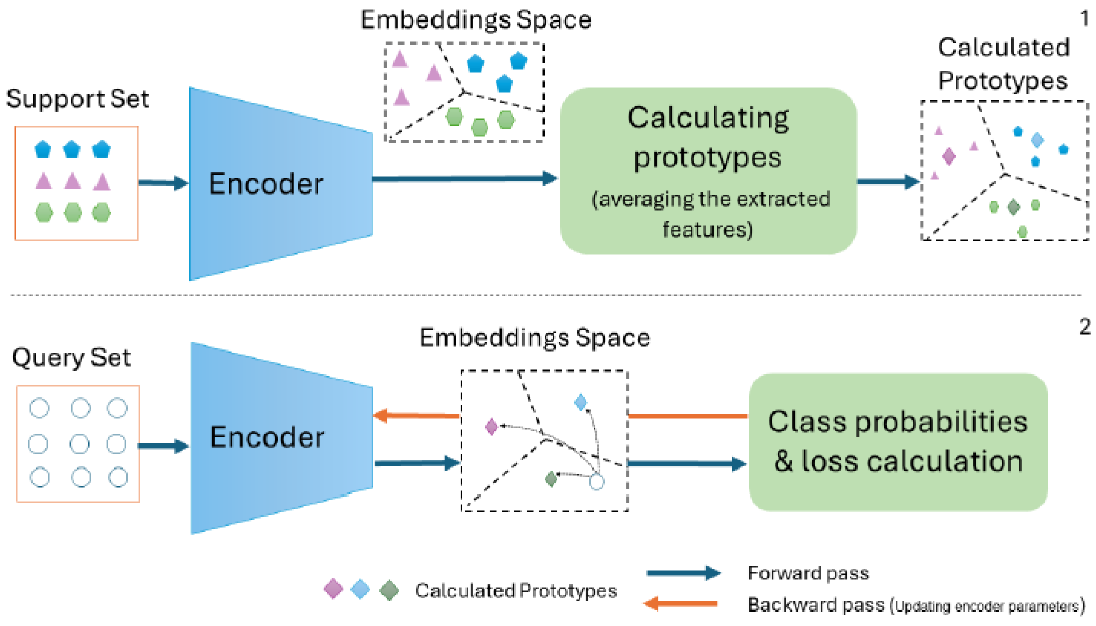
We propose *Prototypical Networks* for the problem of few-shot classification, where a classifier must generalize to new classes not seen in the training set, given only a small number of examples of each new class. Prototypical Networks learn a metric space in which classification can be performed by computing distances to prototype representations of each class. Compared to recent approaches for few-shot learning, they reflect a simpler inductive bias that is beneficial in this limited-data regime, and achieve excellent results. We provide an analysis showing that some simple design decisions can yield substantial improvements over recent approaches involving complicated architectural choices and meta-learning. We further extend Prototypical Networks to zero-shot learning and achieve state-of-the-art results on the CU-Birds dataset.

- Prototypical networks are a type of <u>few-shot learning model</u>
- Few-shot learning is to learn from a limited number of labeled examples (or "shots") per class.

Snell, J., Swersky, K., and Zemel, R., "Prototypical networks for few- shot learning," Advances in neural information processing systems, vol. 30, 2017.

How Prototypical Networks are Trained

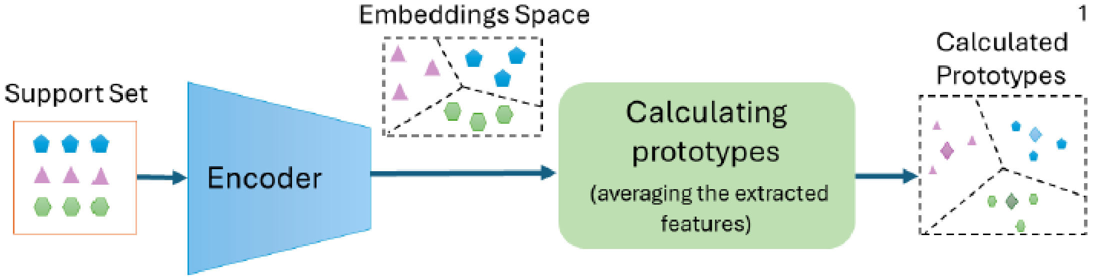
- **Embedding Function**: Each input (e.g., an image) is passed through an embedding function (a neural network) to produce a feature representation.
- **Prototypes**: For each class, the model calculates the prototype by averaging the feature representations of the few labeled examples provided for that class.
- **Distance Metric**: A query set of samples is classified by measuring their distance (typically using Euclidean distance) to the prototypes in the feature space. The sample is assigned to the class whose prototype is closest. The loss is computed and prototypes updated.



Snell, J., Swersky, K., and Zemel, R., "Prototypical networks for few- shot learning," Advances in neural information processing systems, vol. 30, 2017.

Using PNs on new domains after training is simple

- **Feature Extraction**: After training, prototypical networks use the previously learned encoder to map new input data into a lower-dimensional feature space.
- **Prototype Computation**: For each class in the new support set (a few labeled examples per class), the network computes a prototype by averaging the feature embeddings of the examples belonging to that class.
- **Distance-Based Classification**: When a new (query) sample arrives, it is mapped to the same feature space, and its embedding is compared to the class prototypes using a distance metric (typically Euclidean distance). The query sample is assigned to the class with the closest prototype, enabling classification with minimal labeled data.
- **Few-Shot Capability**: If the encoder has extracted representative features, it will generalize well to unseen classes and can classify new categories with only a few examples, making it useful for few-shot learning scenarios.



Prototypical Networks

- Prototypical networks are effective for tasks with scarce labeled data, making them ideal for applications where data collection is challenging or expensive.

Many successful applications

- **Medical Diagnosis**: In medical imaging, prototypical networks help classify diseases when there are only a few examples of certain conditions available.
- **Speech Recognition**: Used for recognizing new speech patterns or accents with limited training data.

Snell, J., Swersky, K., and Zemel, R., "Prototypical networks for few- shot learning," Advances in neural information processing systems, vol. 30, 2017.

SCHULICH
School of Engineering

Can PNs learn good Radio Embeddings?

PNs were proposed for few-shot learning in images – can we use :

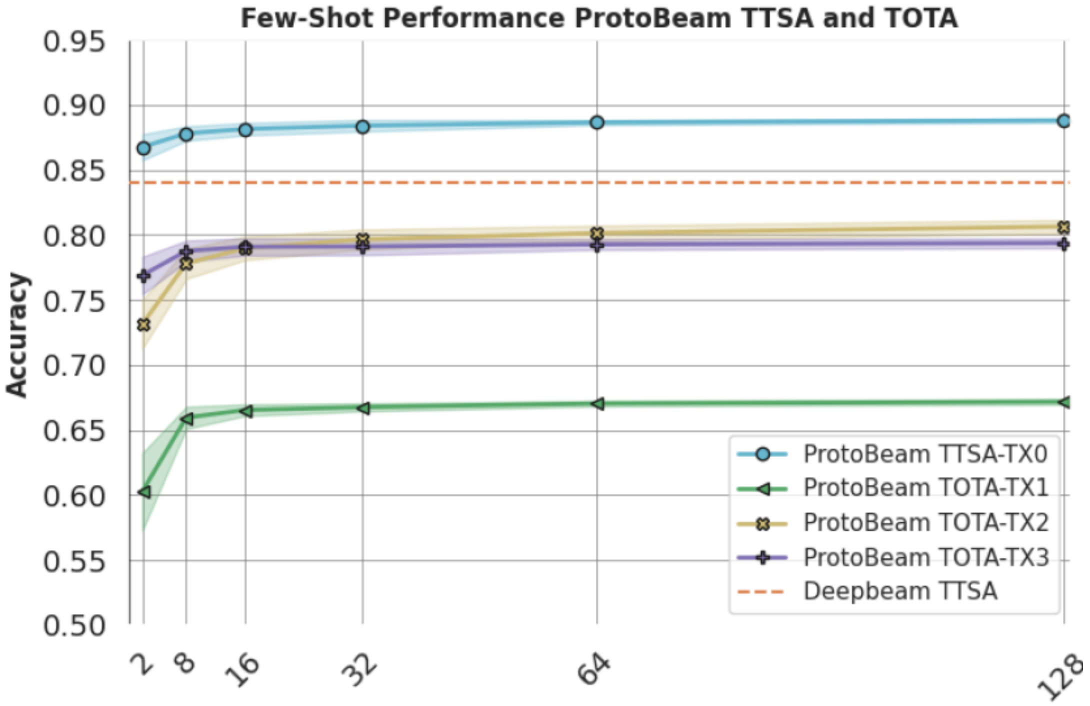
- their property of learning general representations to create ***radio embeddings*** that are robust to wireless domain shifts?
- and the few-shot property to quickly adapt to these domain shifts?

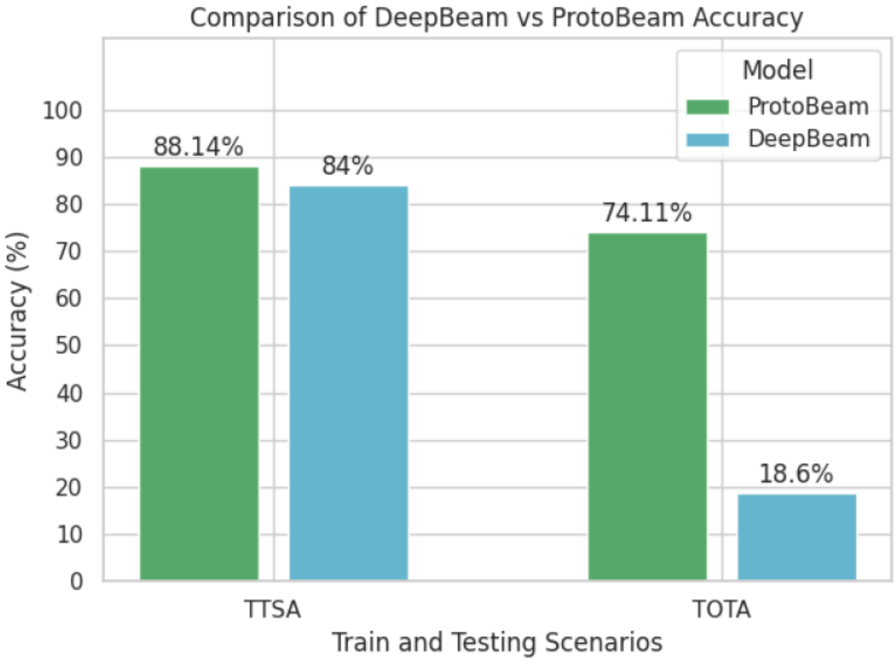
GOAL: Learn generalizable radio embeddings/ representations that can be quickly adapted with a few shots of calibration if the domain changes.

ProtoBeam: *Generalization to Unseen Antennas*

UNIVERSITY OF CALGARY

- ProtoBeam improved the accuracy significantly in the TOTA scenario with only 2-8 samples/beam of labeled data from the different RF front-end



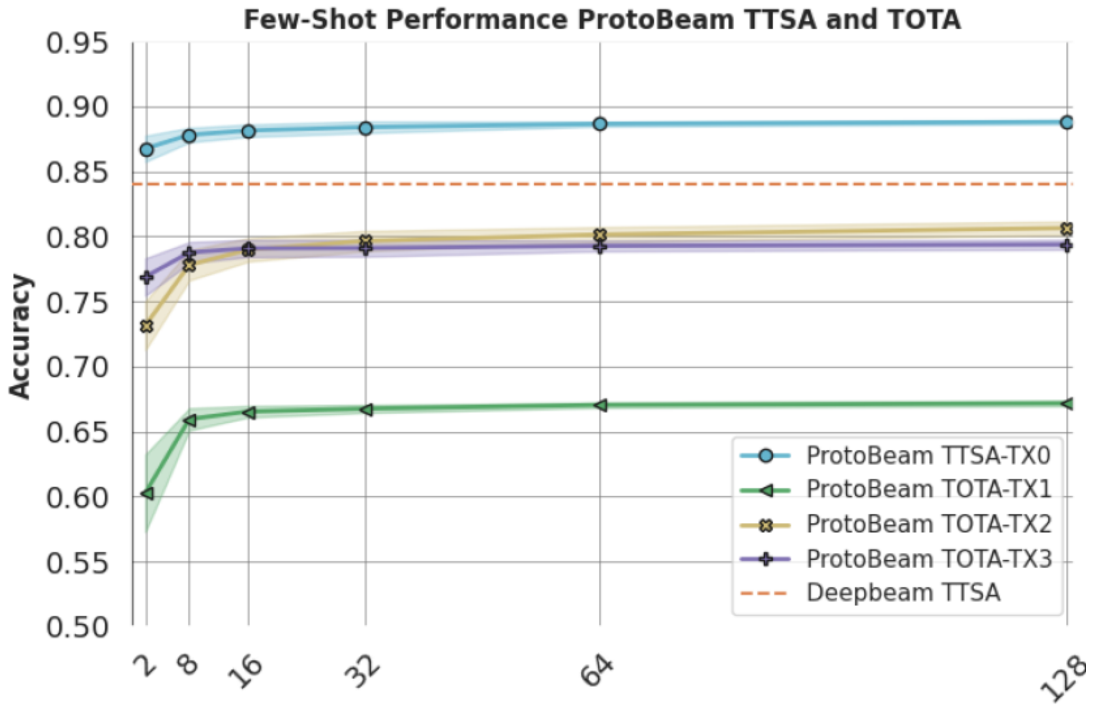


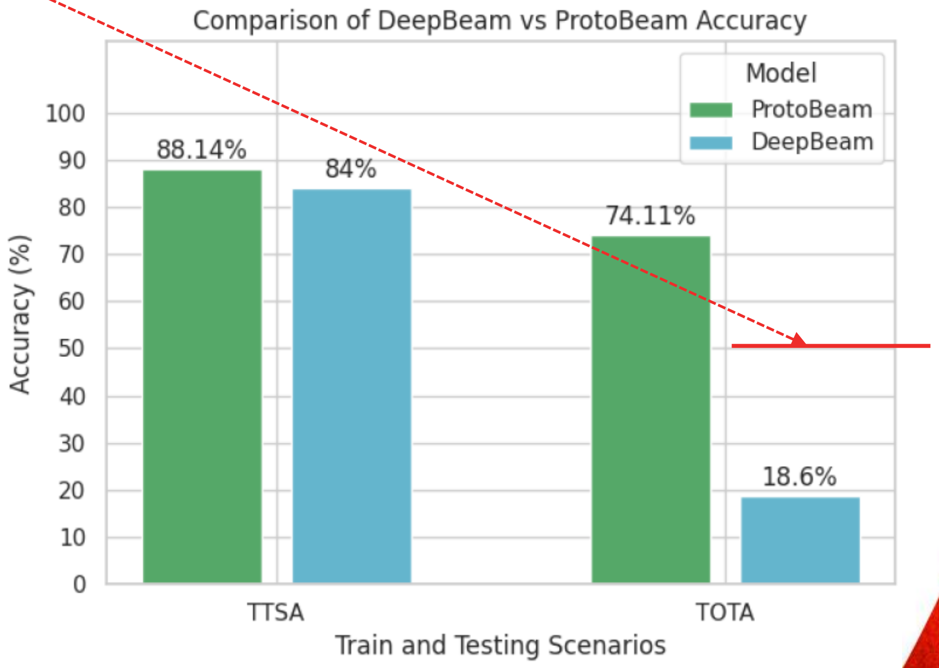
O. Mashaal, et. al. "ProtoBeam: Generalizing Deep Beam Prediction to Unseen Antennas using Prototypical Networks." IEEE GLOBECOM 2024

SCHULICH
School of Engineering

ProtoBeam: *Generalization to Unseen Antennas*

- ProtoBeam performed better than a mixed training setting with all the RF front-end data (50%)
- ProtoBeam also improved the baseline TTSA accuracy





O. Mashaal, et. al. "ProtoBeam: Generalizing Deep Beam Prediction to Unseen Antennas using Prototypical Networks." IEEE GLOBECOM 2024

ProtoBeam Training

Algorithm 1 Proposed ProtoBeam Training Algorithm

Inputs: Training set $\mathcal{D} = \{(x_1, y_1), \ldots, (x_{n_b}, y_{n_b})\}$, where x_i is the I/Q sample and $y_i \in \{1, \ldots, B\}$ is the target beam. $\mathcal{D}_b$ denotes the subset of $\mathcal{D}$ containing all elements (x_i, y_i) for target beam b.
Parameters: n_b is the number of baseband I/Q samples. B is the number of target beams, $N_B \leqslant B$ is the number of target beams per episode.
n_S is the number of I/Q support examples per target beam.
n_Q is the number of I/Q query examples per target beam.
$RandSample(\mathcal{S}, N)$ denotes a set of N elements chosen uniformly at random from set $\mathcal{S}$, without replacement.
Output: Updated model parameters after backpropagation.

```
procedure TrainProtoBeam (D)
    Select indices for target beams in this episode
    V ← RandSample({1, ..., B}, N_B)
    for b ∈ V do
        I_b ← RandSample(D_b, n_S)                          ▷ Support
        Q_b ← RandSample(D_b \ I_b, n_Q)                    ▷ Query
        p_b ← (1/n_S) Σ_{(x_i,y_i)∈I_b} f_φ(x_i)            ▷ Compute Prototypes
    end for
    L ← 0                                   ▷ Initialize loss for this episode
    for b ∈ V do
        for (x_i, y_i) ∈ Q_b do
            L ← L + 1/(N_B n_Q) [ d(f_φ(x_i), p_b) +
                 log( Σ_{b'} exp( − d(f_φ(x_i), p_{b'}) ) ) ]
        end for
    end for
    Compute gradients of L w.r.t. model parameters φ
    Perform backpropagation to update model parameters
    φ ← φ − α · ∇_φ L                      ▷ Update φ with learning rate α
    return φ
end procedure
```

$$p_b \leftarrow \frac{1}{n_S} \sum_{(x_i, y_i) \in I_b} f_\phi(x_i)$$

$$L \leftarrow L + \frac{1}{N_B n_Q}\left[d(f_\phi(x_i), p_b) + \log\left(\sum_{b'} \exp\left(- d(f_\phi(x_i), p_{b'}) \right) \right) \right]$$

$$\phi \leftarrow \phi - \alpha \cdot \nabla_\phi L$$

O. Mashaal, et. al. "ProtoBeam: Generalizing Deep Beam Prediction to Unseen Antennas using Prototypical Networks." IEEE GLOBECOM 2024

Enhancing ProtoBeam with Augmentations

- We tested normalization, data augmentation, and prototype normalization on model accuracy. Starting from a 38.67% baseline against different antenna configurations, each technique was applied sequentially

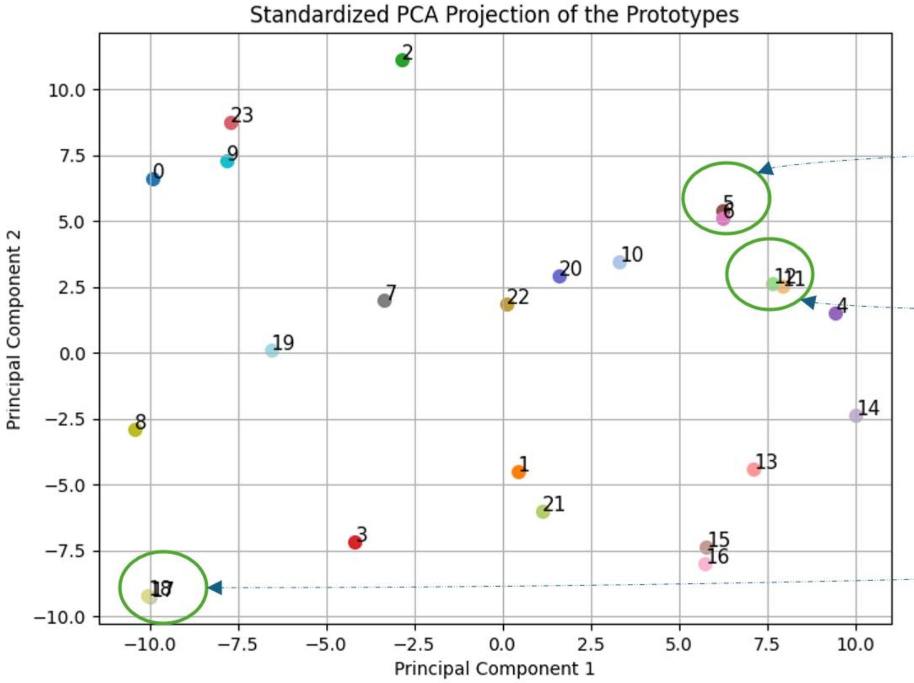
TABLE I: ProtoBeam Performance with Data Augmentation and Normalization.

Experimental Setup	TTSA (%)		TOTA (%)	
	2-shot	32-shot	2-shot	32-shot
w/o Data Norm. or Augm.	61.2	73.4	38.67	42.8
Data Normalization only	77.3	83.3	45.8	56.8
Data Norm. & Augm.	79.69	83.68	49.9	60.4
Prototypes Norm. + Data Norm & Augm.	81.9	84.5	55.26	64.2

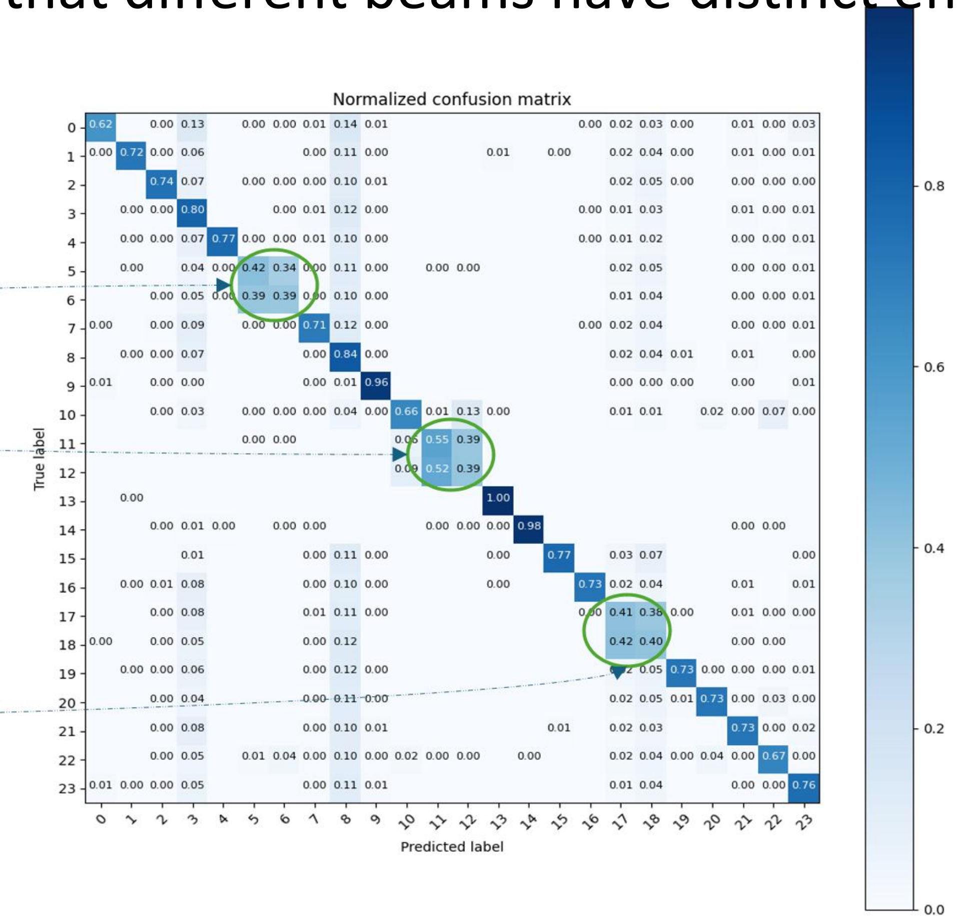
O. Mashaal, et. al. "ProtoBeam: Generalizing Deep Beam Prediction to Unseen Antennas using Prototypical Networks." IEEE GLOBECOM 2024

Interpreting the Embedding Space via PCA

- ProtoBeams mistakes were mainly in adjacent beams
- PCA projections of the embeddings show that different beams have distinct embeddings



PCA projections



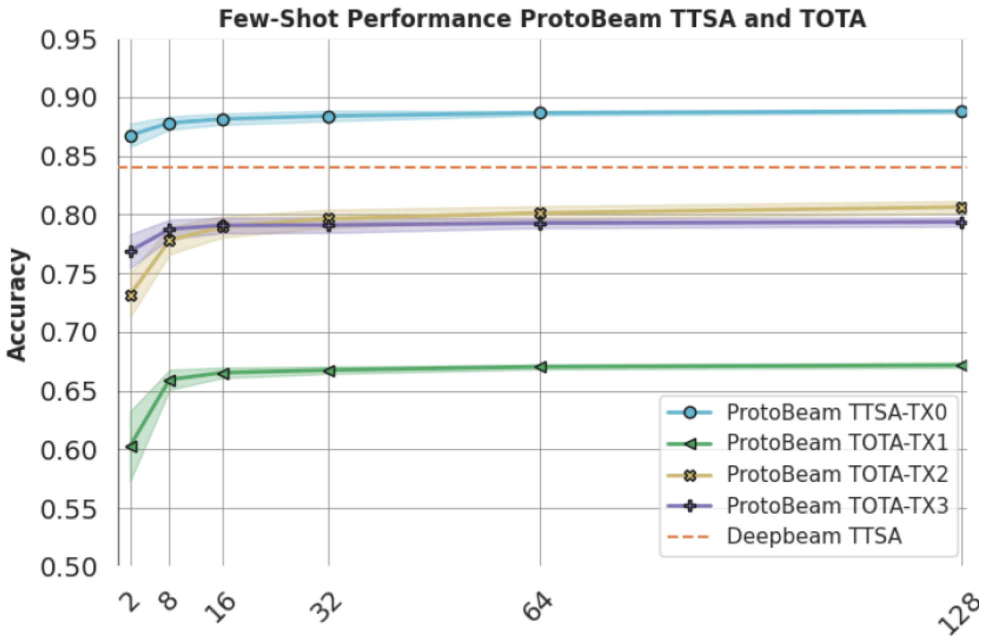
O. Mashaal, et. al. "ProtoBeam: Generalizing Deep Beam Prediction to Unseen Antennas using Prototypical Networks." IEEE GLOBECOM 2024

Key Takeaway [1/2]

Protobeam provided an

AI model that can generalize to multiple RF front ends by

Further tuning with only 2-8 labeled samples per beam only produced promising results

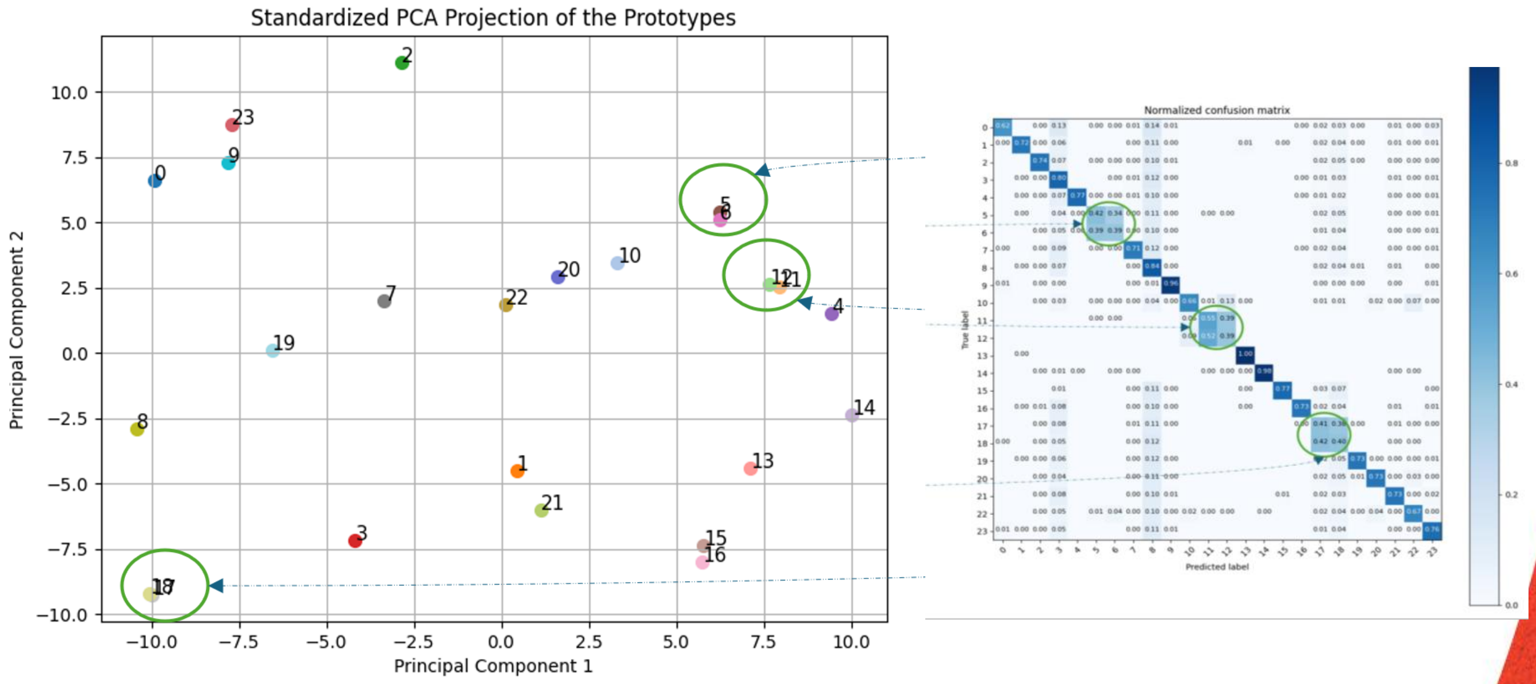


A. Aboulfotouh, et. al. "ProtoBeam: Generalizing Deep Beam Prediction to Unseen Antennas using Prototypical Networks." IEEE GLOBECOM 2024

Key Takeaways [2/2]

Protobeam provided

representations that are interpretable in lower feature spaces



A. Aboulfotouh, et. al. "ProtoBeam: Generalizing Deep Beam Prediction to Unseen Antennas using Prototypical Networks." IEEE GLOBECOM 2024

PNs to reduce data collection campaigns

ProtoAoA: Protoypical Networks for Few-Shot Angle-of-Arrival Estimation

Elsayed Mohammed‡, Omar Mashaal‡, Alec Digby*, Ashkan Eshaghbeigi*, Hatem Abou-Zeid‡
‡Department of Electrical and Software Engineering, University of Calgary, Canada
*Qoherent Inc., Toronto, Ontario, Canada

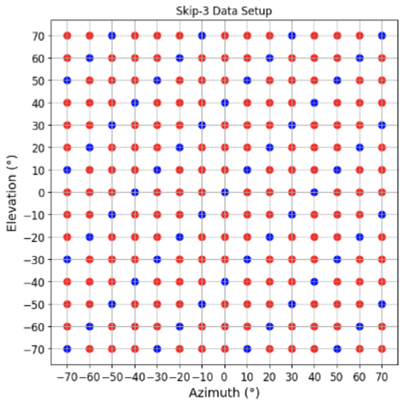


Fig. 2: Proposed "Skip-3" few shot learning AoA data set-up.

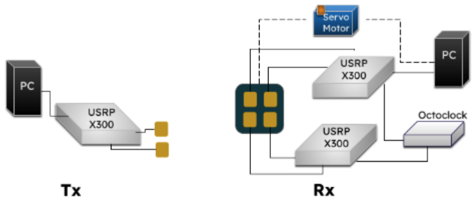


Fig. 3: A diagram of the testbed used to create the dataset used through our study.

Building Radio Foundation Models using Transformers

Ahmed Aboulfotouh
PhD student

Qoherent
mitacs
NSERC CRSNG
UNIVERSITY OF CALGARY

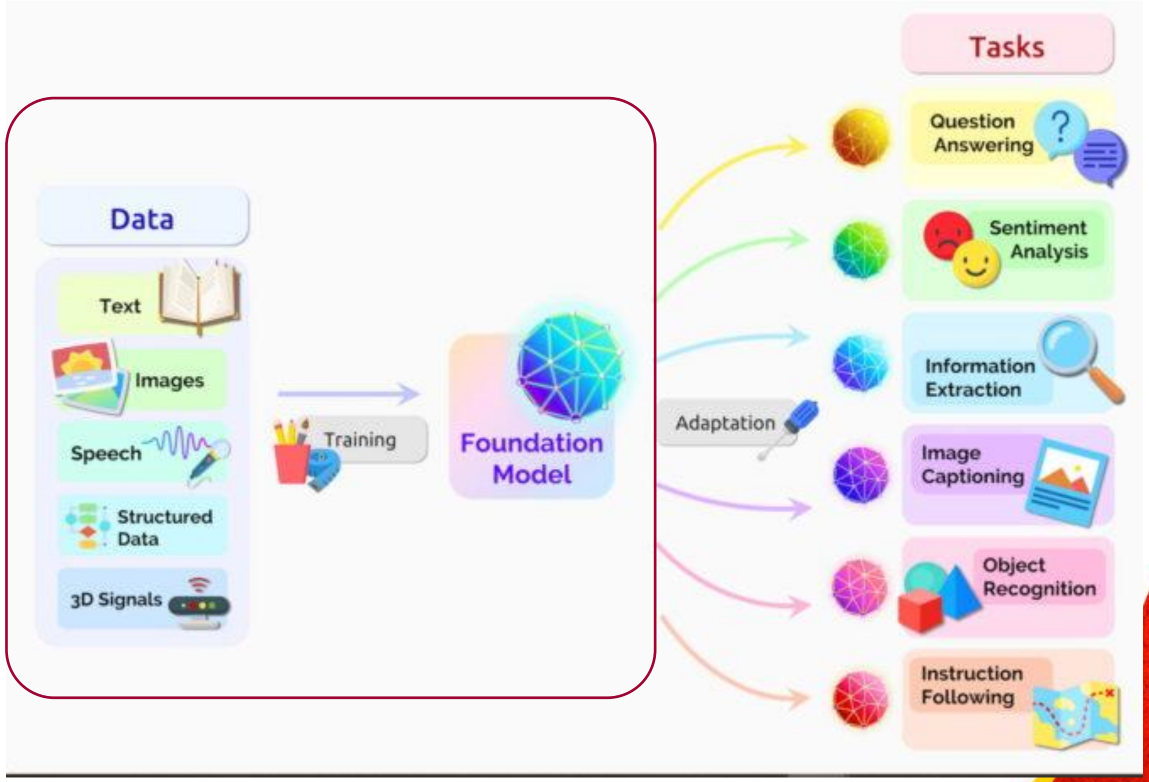
Can we do better than Domain Adaptation?

1. Can we learn more generic representations from wireless signals for multiple-tasks?
   - Generalizable Radio Embeddings that enable multiple downstream tasks?

2. Can we learn these embeddings in a self-supervised fashion without requiring labeled data?

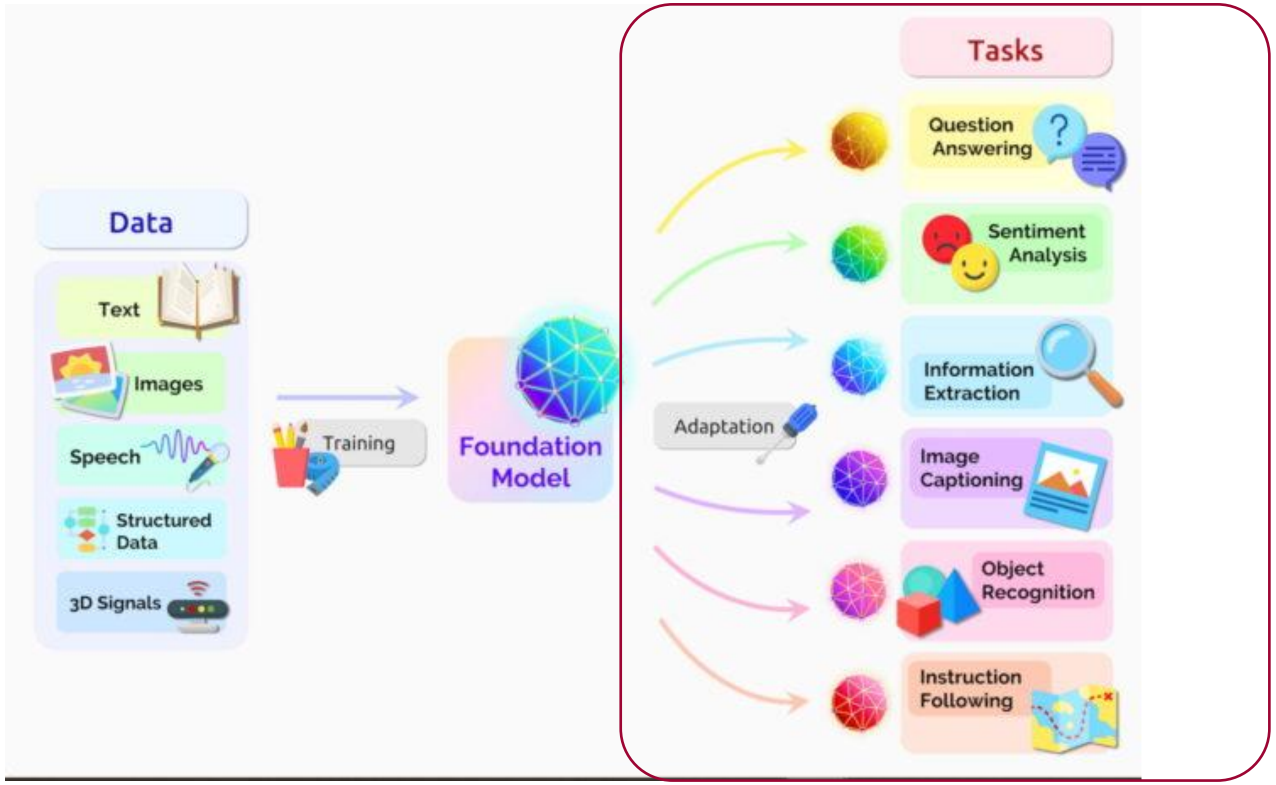
How are Foundation Models trained?

- Typically, on large datasets using self-supervised learning (SSL)
- SSL does not rely on manually labeled data and uses the structure of the data itself to create "learning objectives".



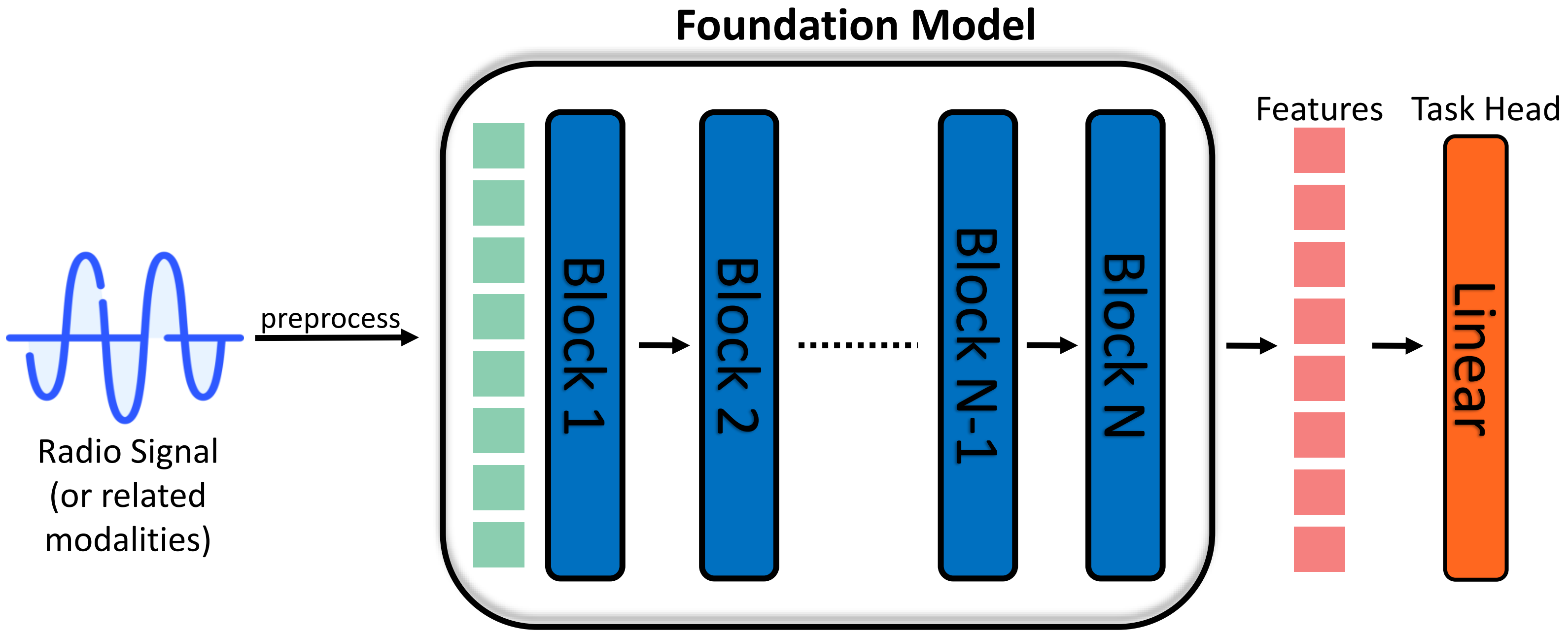
https://blogs.nvidia.com/blog/what-are-foundation-models/

Foundation Models (FM): *Fine-tuning*



https://blogs.nvidia.com/blog/what-are-foundation-models/

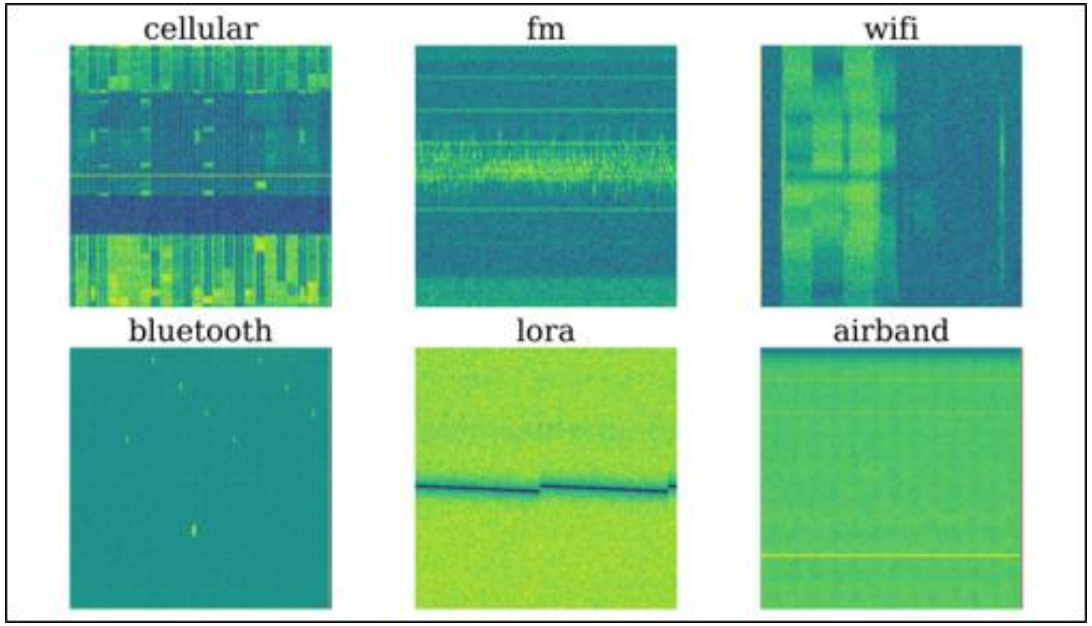
(How to) Build Radio Foundation Models



- The foundation model's output features serve as inputs to task-specific heads.
- We primarily fine-tune a task head for each task, but some foundation model blocks can also be unfrozen for further tuning.
- **One model, multiple tasks**: By swapping task heads, we can use the same model for different applications.

Task-1: Radio Signal Identification

- The dataset for this task consists of spectrograms representing various types of radio signals, categorized into 20 classes.

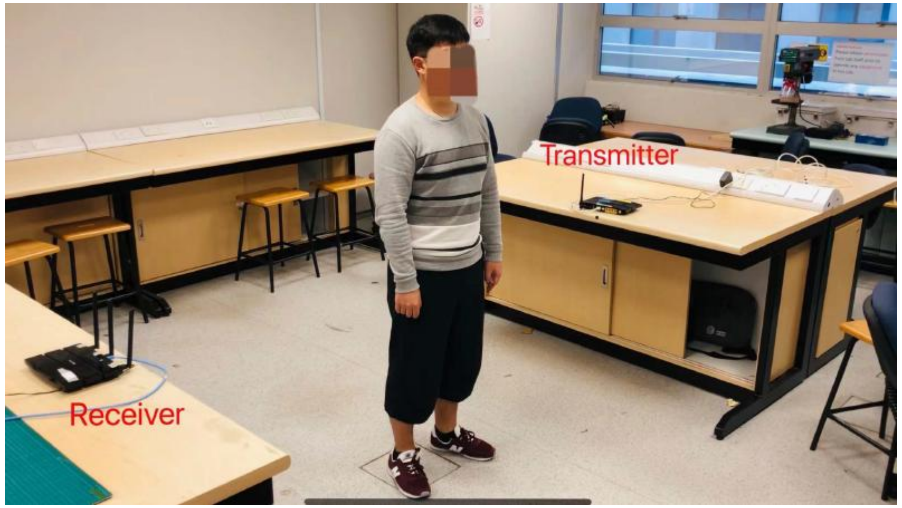


Sample from the dataset

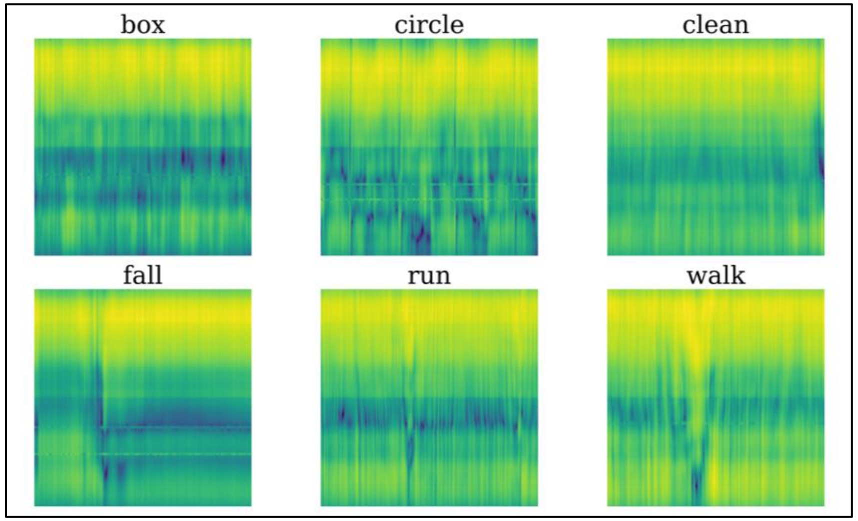
CommRad RF: A dataset of communication radio signals for detection, identification and classification

Task-2: WiFi Human Activity Sensing

- The dataset contains channel state information (CSI) measurements for six human activities: running, walking, falling, boxing, arm circling, and floor cleaning.
- Participants performs these activities between a pair of Wi-Fi access points, and CSI is measured for each activity.



Data Collection

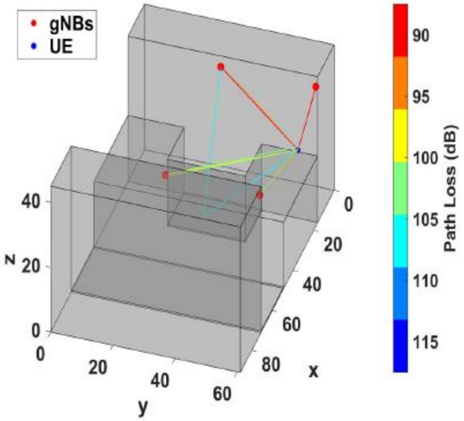


Sample from the dataset

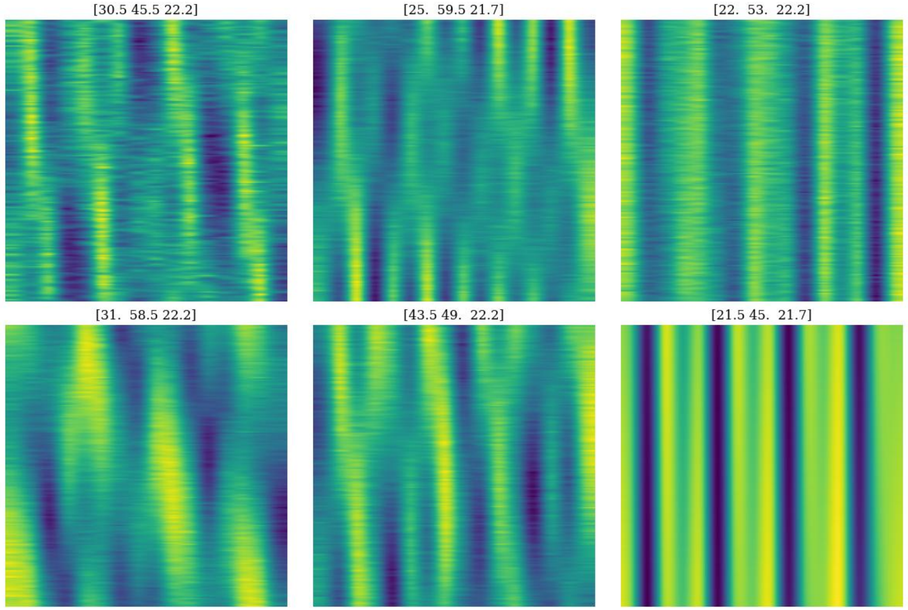
Yang, Jianfei, et al. "EfficientFi: Toward large-scale lightweight WiFi sensing via CSI compression." IEEE Internet of Things Journal 9.15 (2022): 13086-13095.

Task-3: 5G Positioning

- This dataset contains CSI measurements from reference signals exchanged between a user and four base stations. The task is to predict the UE's position based on its CSI.



Data Collection



Sample from the dataset

Gao, Kaixuan, Huiqiang Wang, and Hongwu Lv. "CSI dataset towards 5G NR high-precision positioning." *IEEE Dataport* (2021).

Recap of Goals

1. Can we learn generic representations from wireless signals for multiple tasks?

   Demonstrate the effectiveness of a Vision Transformer encoder to achieve this

2. Can we learn these embeddings in a self-supervised fashion <u>without requiring labeled data</u>?

   Propose: Masked Spectrogram Modeling and use a real-world unseen unlabeled dataset for training

Building 6G Radio Foundation Models with Transformer Architectures

Ahmed Aboulfotouh[‡], Ashkan Eshaghbeigi[*], and Hatem Abou-Zeid[‡]
[‡]Department of Electrical and Software Engineering, University of Calgary, Canada
[*]Qoherent Inc., Toronto, Ontario, Canada

Abstract—**Foundation deep learning (DL) models are general models, designed to learn general, robust and adaptable representations of their target modality, enabling finetuning across a range of downstream tasks. These models are pretrained on large, unlabeled datasets using self-supervised learning (SSL). Foundation models have demonstrated better generalization than traditional supervised approaches, a critical requirement for wireless communications where the dynamic environment demands model adaptability. In this work, we propose and demonstrate the effectiveness of a Vision Transformer (ViT) as a *radio foundation model* for spectrogram learning. We introduce a Masked Spectrogram Modeling (MSM) approach to pretrain the ViT in a self-supervised fashion. We evaluate the ViT-based foundation model on two downstream tasks: Channel State Information (CSI)-based Human Activity sensing and Spectrogram Segmentation. Experimental results demonstrate competitive performance to supervised training while generalizing across diverse domains. Notably, the pretrained ViT model outperforms a four-times larger model that is trained from scratch on the spectrogram segmentation task, while requiring significantly less training time, and achieves competitive performance on the CSI-based human activity sensing task. This work demonstrates the effectiveness of ViT with MSM for pretraining as a promising technique for scalable foundation model development in future 6G networks.**

Index Terms—**Self-Supervised Learning, Foundation Models, Deep Learning, Human Activity Sensing, Spectrogram Segmentation**

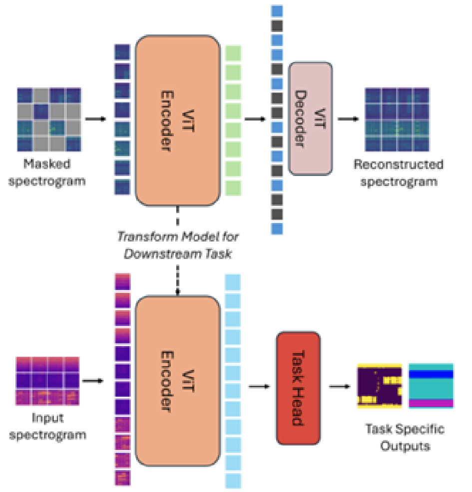


Fig. 1: Proposed ViT Foundation Model for Radio Spectrograms.

I. Introduction

Foundation models (FMs) are first trained on a large, often unlabeled dataset, allowing them to build broad, adaptable representations that can be finetuned for various downstream tasks. This initial pretraining stage is done using self-supervised learning (SSL), where the model learns underlying patterns and relationships within the data without relying on labeled examples [1]–[3]. The model ideally develops a robust understanding of its target modality, which, in our case, is radio spectrograms.

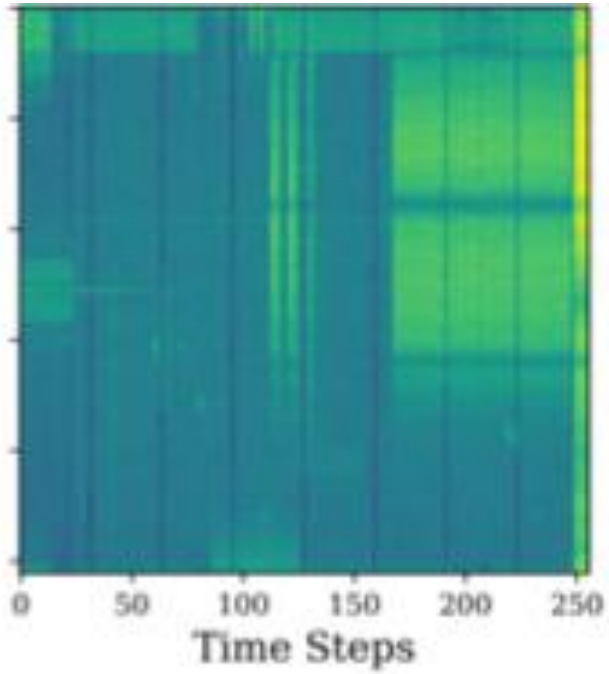
In fields like computer vision and natural language processing, FMs have set new benchmarks [4]–[7], often surpassing supervised learning models, specifically designed for individual tasks. This is largely due to their ability to generalize: FMs learn flexible and transferable representations that make them better suited to handle variations in data, perform across diverse tasks, and adapt to new contexts. Generalization is especially

can perform well with minimal additional labeled samples.] Deep learning (DL) has demonstrated strong potential when applied to individual wireless tasks, including automatic modulation classification [8], channel estimation [9], constellation and waveform design [10], among others. However, these models are highly specialized, and there are concerns about their ability to generalize effectively in real-world scenarios. Wireless signals are subject to time-varying impairments, and the communication environment is constantly changing, which can degrade a DL model's performance if it fails to adapt. Introducing the concept of FMs for wireless can potentially overcome these limitations [11], [12].

We propose FMs for wireless signals as a solution to address

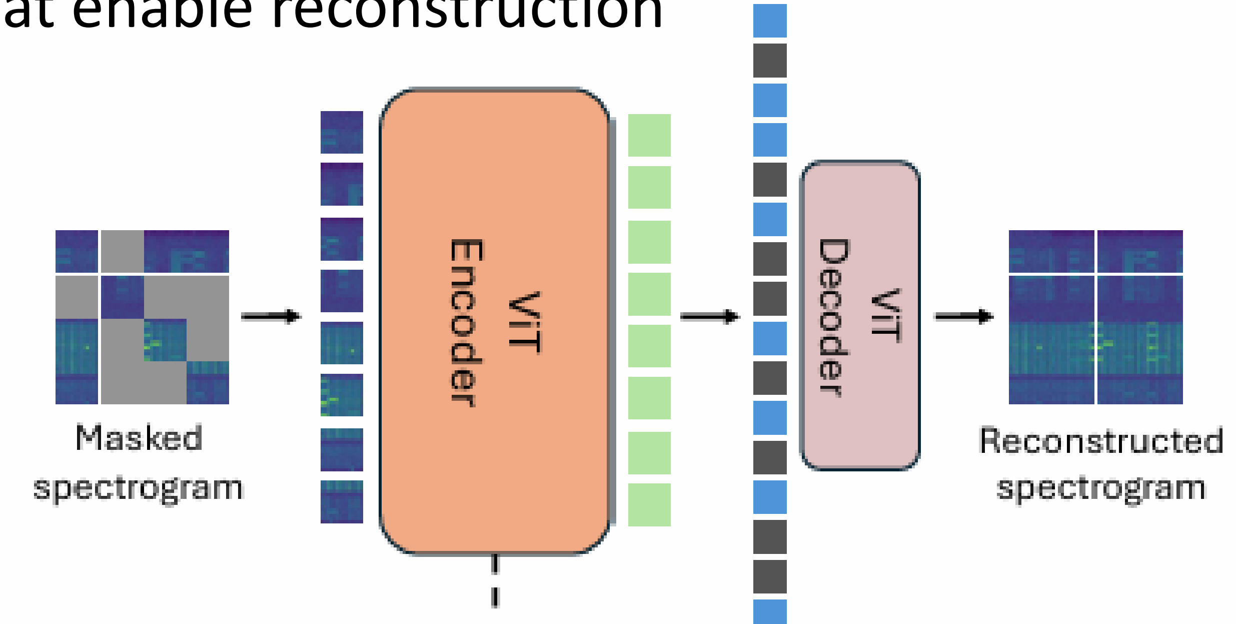
The Dataset used to Build the Foundational Model

- Time-domain recordings of IQ samples in the frequency range of 2.4 to 2.65 GHz, with BW between 10 MHz and 60 MHz using Pluto and Ettus SDRs.
- Samples were collected in downtown Toronto.
- The samples were converted to spectrograms and used for foundational model pretraining



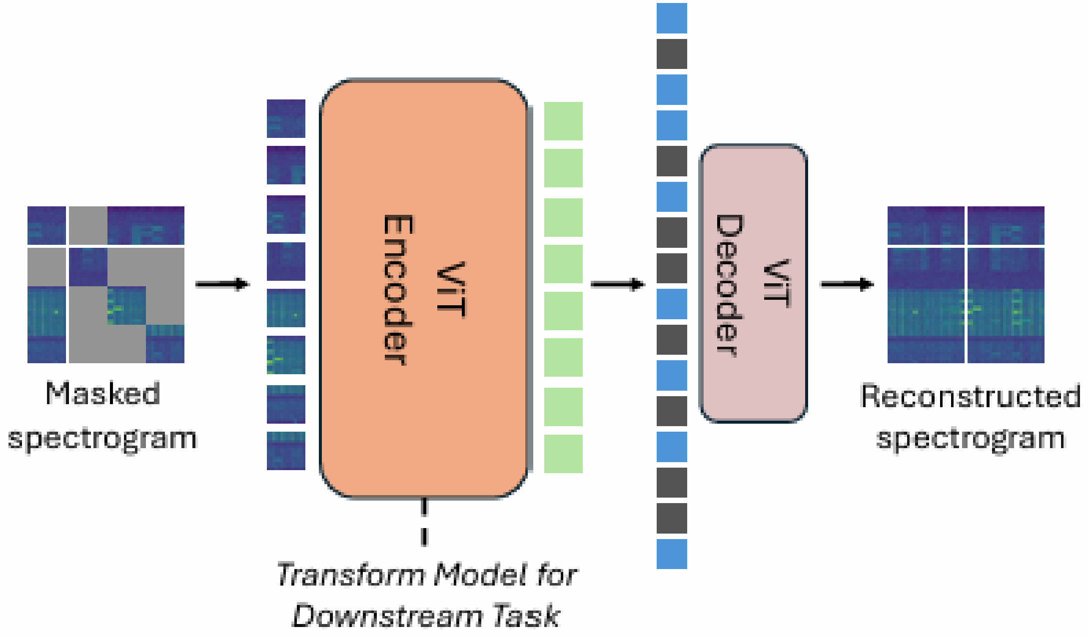
Pre-training with “Masked Spectrogram Modeling”

- The spectrogram is divided into patches and a large %age of the patches is “masked”
- The ViT model’s decoder objective is to reconstruct the original spectrogram from the masked version.
- The model analyzes the surrounding context and infers what was likely in the masked positions. This process is self-supervised as it does not require labels.
- During this process, the goal of the ViT encoder is to learn good representations of the spectrogram that enable reconstruction



Pre-training with Masked Spectrogram Modeling

- <u>Effect of Masking Percentage on Pre-training</u>



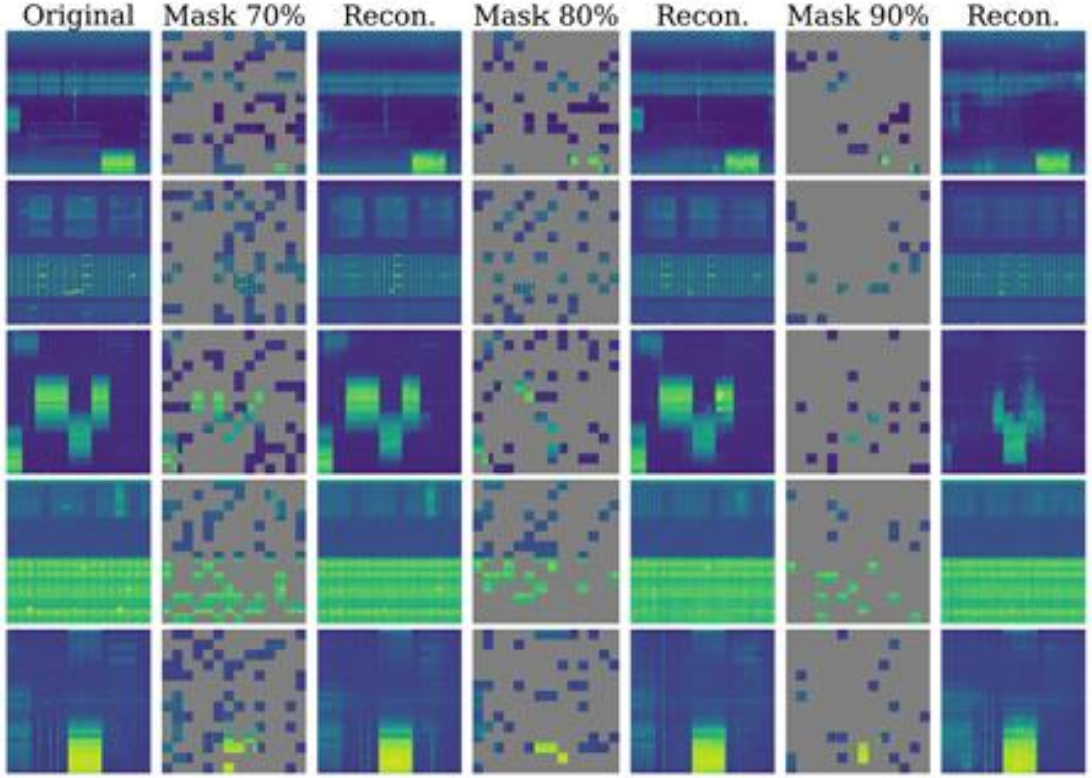


Fig. 4: Reconstruction results of ViT-M at various masking ratios pretrained with a 75% masking ratio.

Pre-training with Masked Spectrogram Modeling

- <u>Effect of Masking Percentage on Pre-training</u>

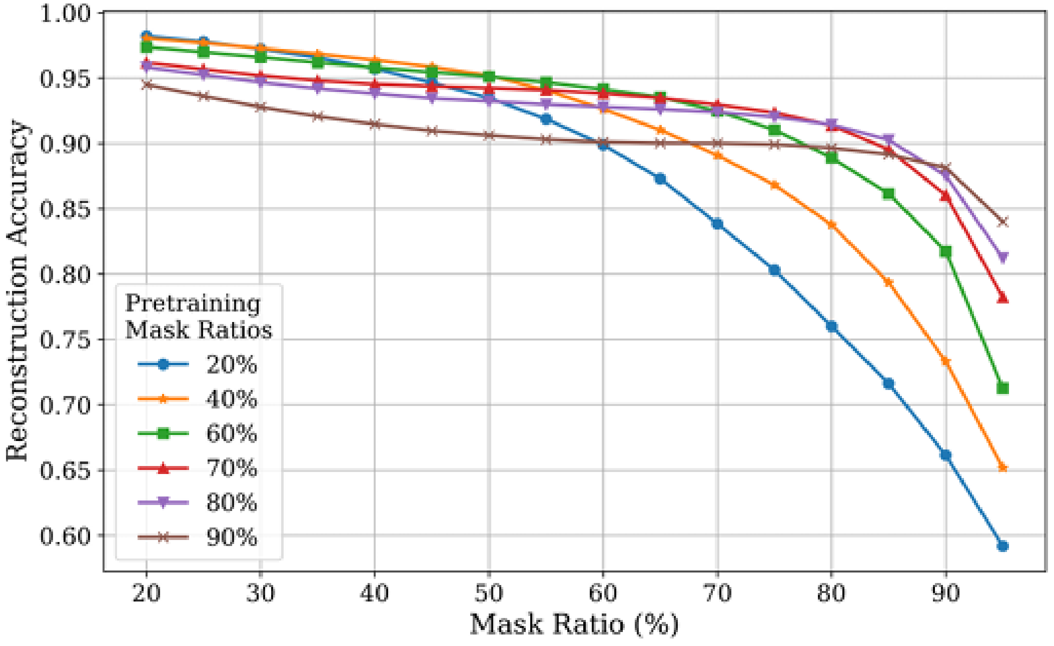
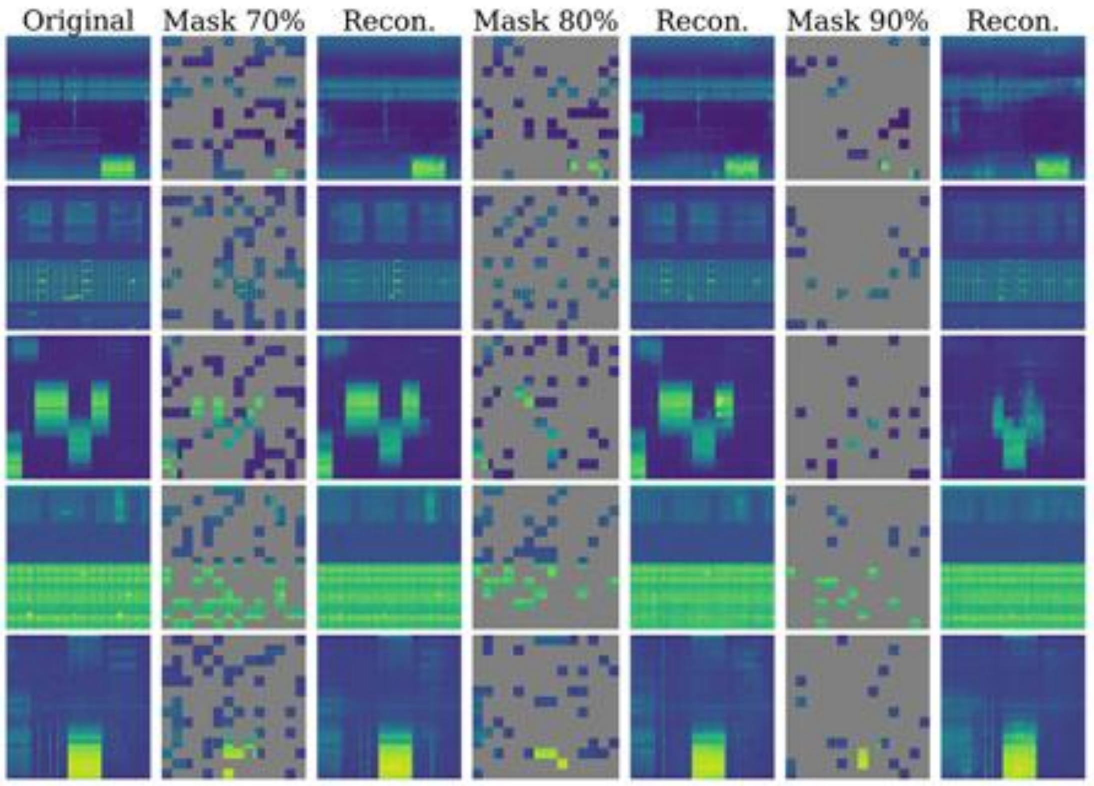


Fig. 6: Reconstruction Accuracy vs Mask ratio of ViT-S pre-trained at various masking ratios.



g. 4: Reconstruction results of ViT-M at various masking ratios pretrained with a 75% masking ratio.

Fine-tuning for Downstream Tasks

- The ViT Decoder is now replaced with a simple multi-layer perceptron (fully connected) neural network
- The new "Task" data is processed through the pretrained ViT Encoder that provides the embeddings/features to the new task head
- Fine-tuning involves learning the task head weights and optionally fine-tuning parts of the ViT encoder

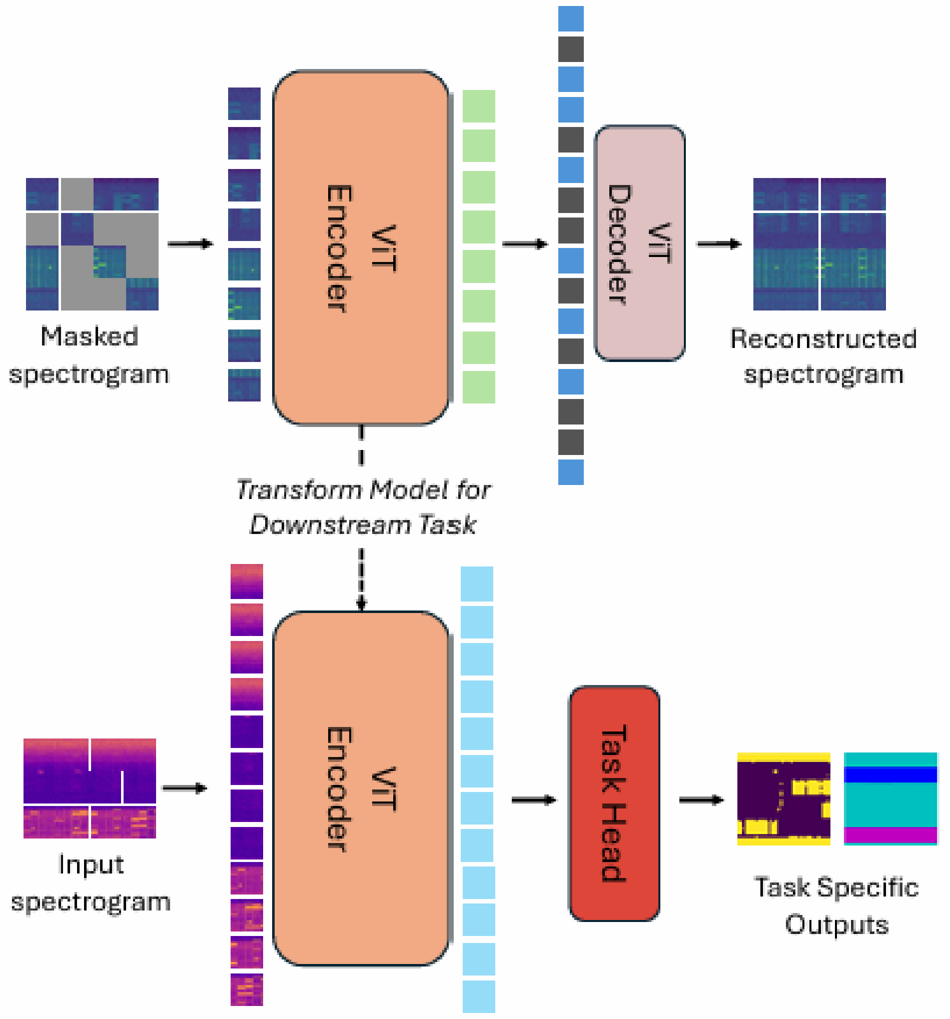


Fig. 3: Proposed ViT Foundation Model for Radio Spectrograms.

A. Aboulfotouh, et. al. "Building 6G Radio Foundation Models with Transformer Architectures." IEEE ICC 2025.

Results for Task-1: Human Activity Recognition

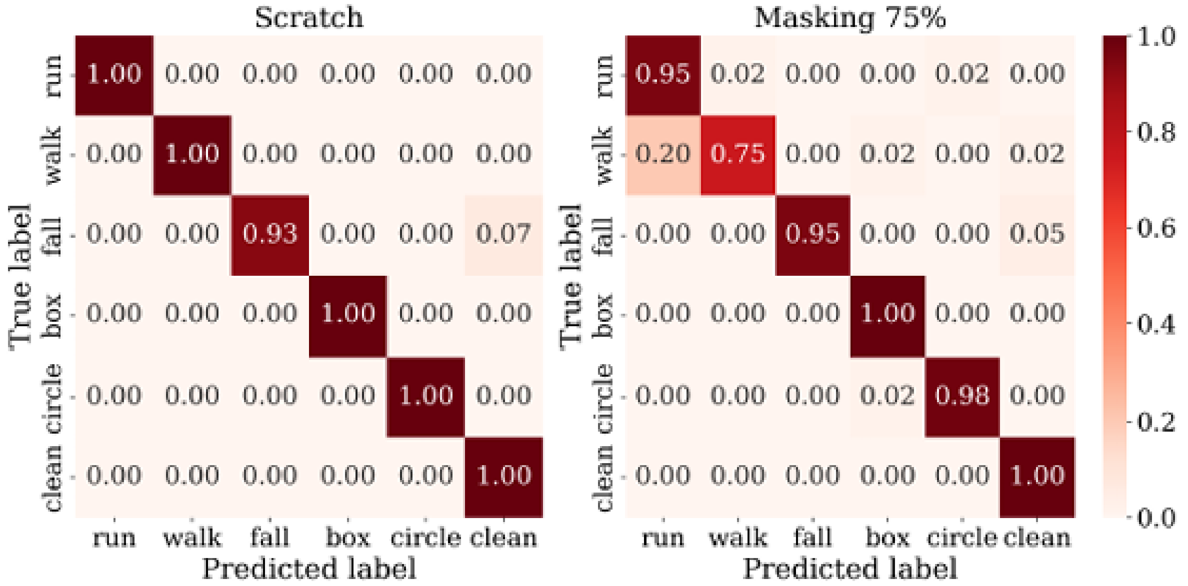
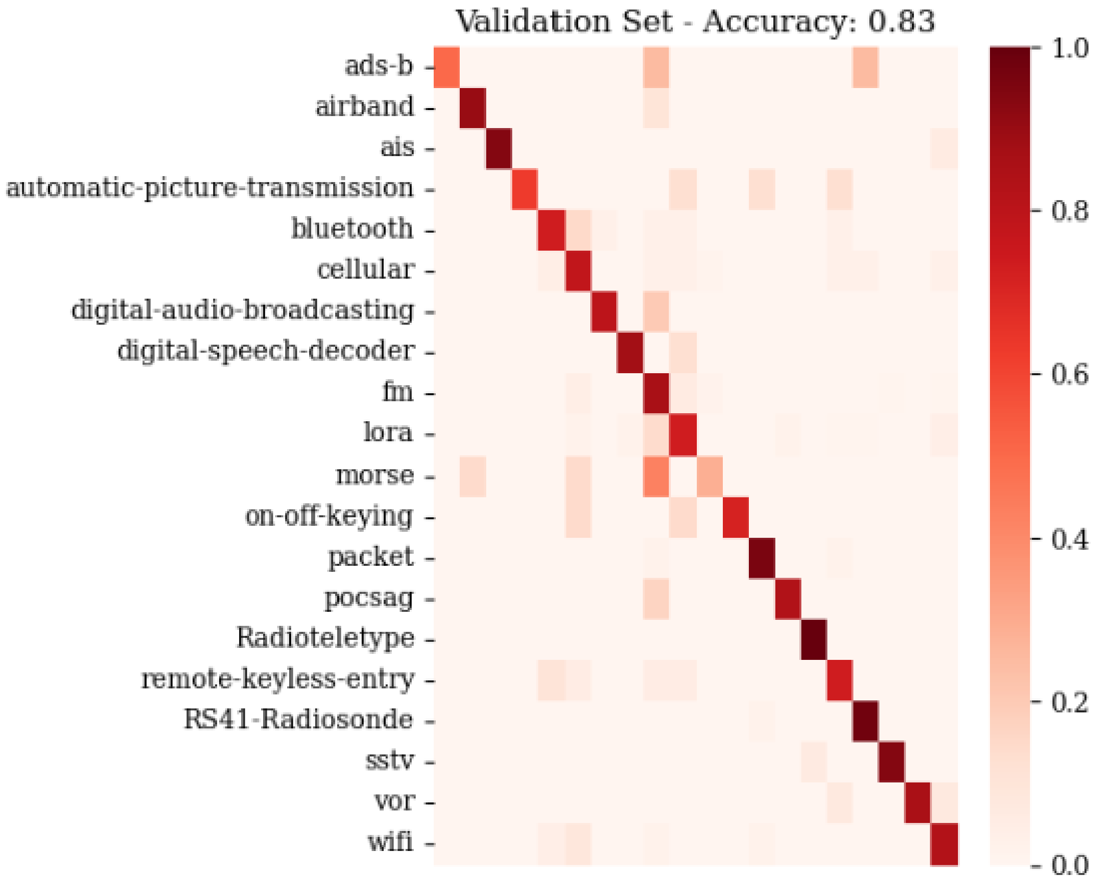


Fig. 7: Confusion matrices of ViT-M trained from scratch and pretrained with a 75% masking ratio.

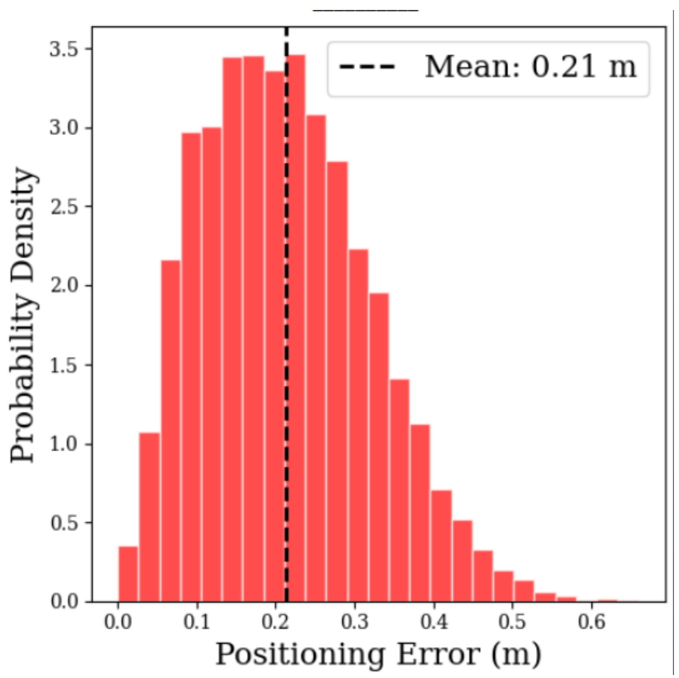
- Results show very close performance to the baselines models that are trained from scratch.
- Important to note that the backbone weights are frozen and only a single linear layer was trained.

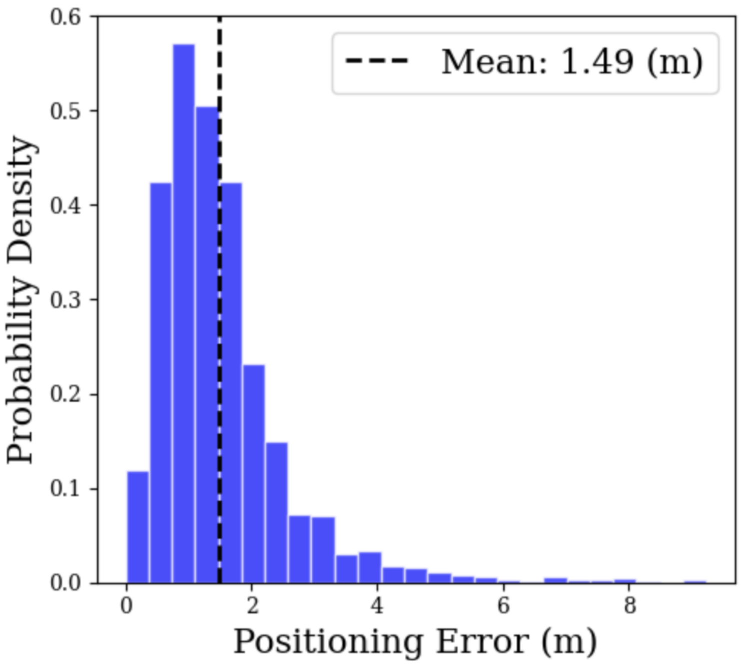
Results for Task-2: RF Signal Classification



- A 2 layer linear layer was trained (single layer was not suffient).

Results for Task-3: 5G CSI Positioning (regression task)

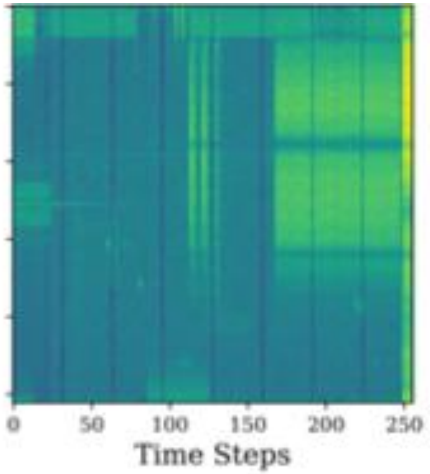




- Despite being a more challenging regression task, the model was able to perform reasonably well
- Linear probing was not sufficient, fine-turning of the last 2 transformer blocks was needed.
- Room for improvement in the test accuracy (FM pre-training data can improve this)

Key Takeaways [1/2]

- The ‘Masked Spectrogram Modeling’ pre-training approach was able to learn radio embeddings with a <u>relatively small dataset of unlabeled spectrogram data</u>.



- The model was successfully fine-tuned to multiple different spectrogram tasks

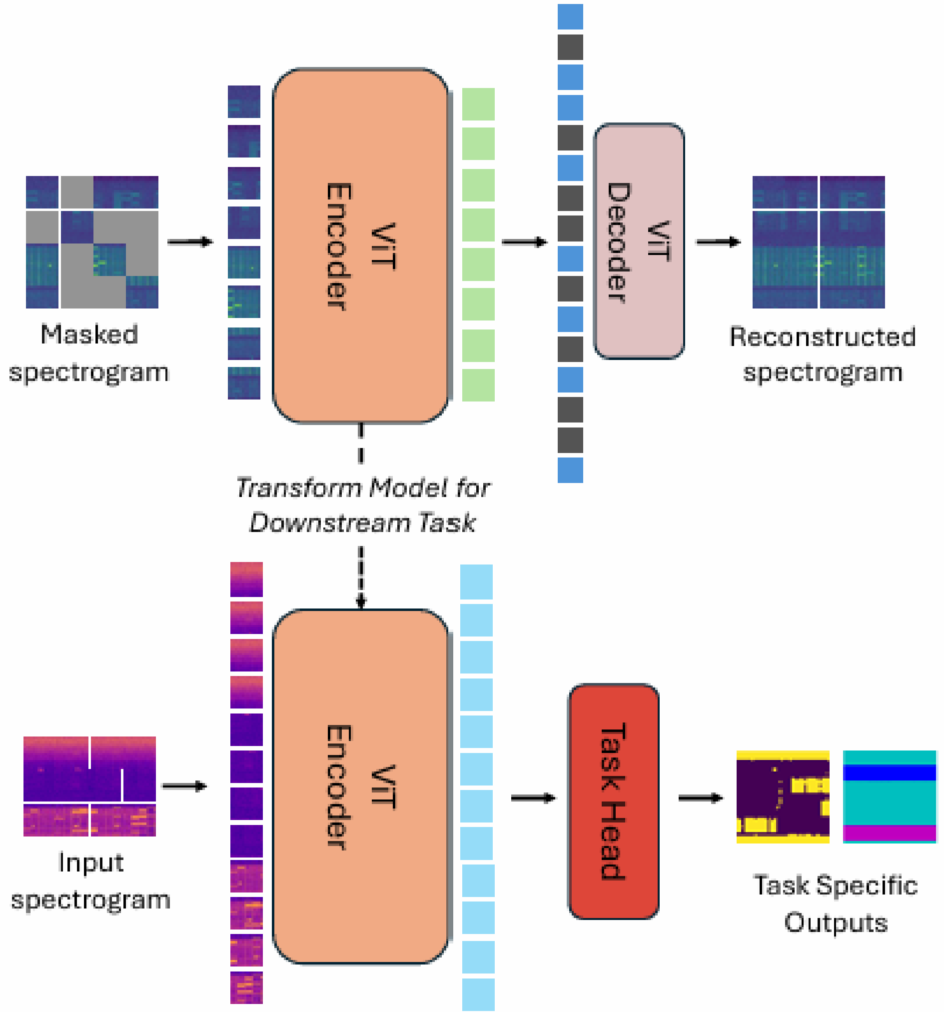


Fig. 3: Proposed ViT Foundation Model for Radio Spectrograms.

A. Aboulfotouh, et. al. "Self-Supervised Radio Pre-training: Toward Foundation Models for Spectrogram L

Key Takeaways [2/2]

- The 'Masked Spectrogram Modeling' pre-training approach was able to learn radio embeddings with a relatively small dataset of unlabeled spectrogram data.
- The model was successfully fine-tuned to two different spectrogram tasks
- This is a first step! More general multi-modal data, pre-training, and fine-tuning approaches need to be investigated.
- Other backbone architectures, radio tokenization and learning effective radio embedding are needed.

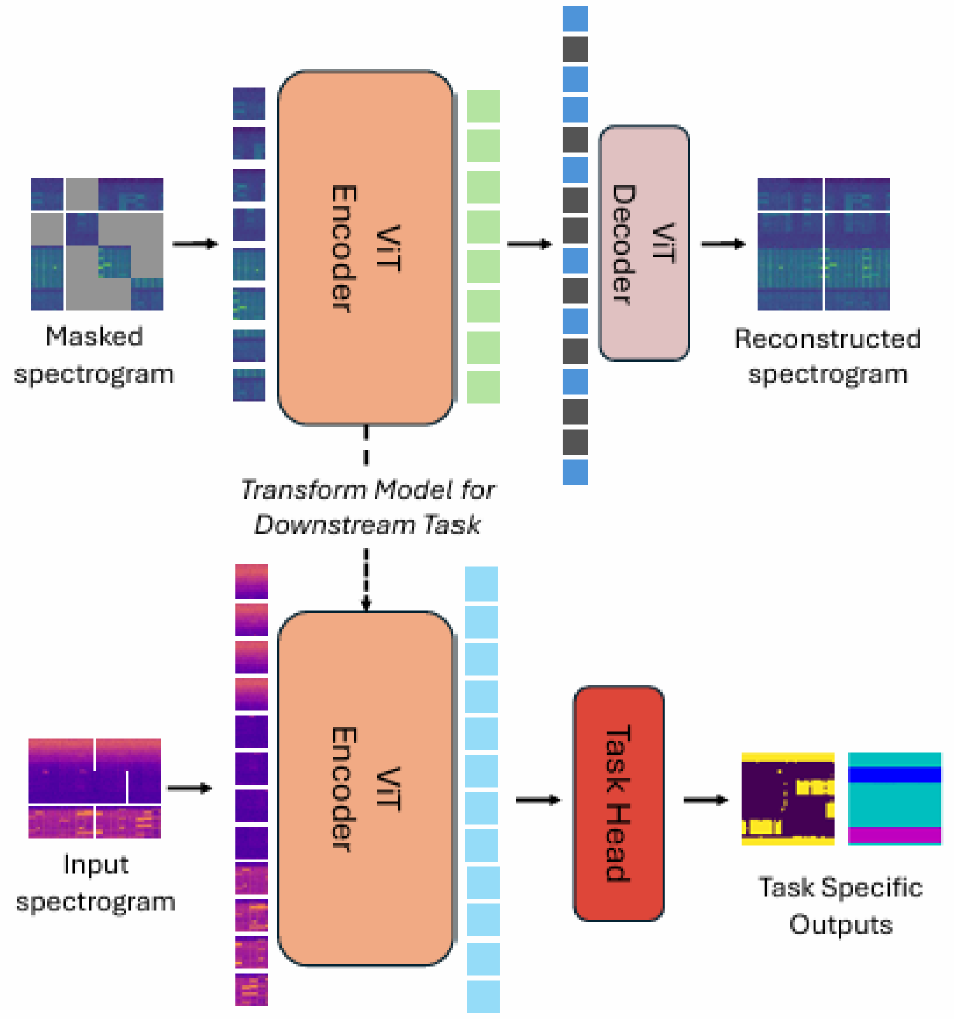


Fig. 3: Proposed ViT Foundation Model for Radio Spectrograms.

Summary & Future Research

Summary

✓ We discussed key AI challenges in wireless networks and the vision of **multi-task radio foundation models** in 6G.

✓ We presented an application of prototypical networks to enable "**generalizable AI**" for Beam Prediction that can domain adapt to different RF-front ends.

✓ We presented an example of building a **spectrogram radio foundation model** that uses a masked spectrogram pre-training approach without requiring labeled data and showed it can be effectively fine-tuned it to multiple tasks.

Opportunities & Future Research

UNIVERSITY OF CALGARY

- Many different functions at the different network layers
  - What constitutes a foundational model in wireless?
- Structure of the wireless data is very different at different layers
  - Multi-modal foundation models?
- Wireless signals are different from language/text and vision where much of the FM research and progress has been made
  - What are good radio data augmentation and label-free pre-training techniques for wireless?
  - What are good base models or architectures for different tasks?

SCHULICH
School of Engineering

How do we get there?

- A lot of small steps are likely needed before full fledged "radio foundational models".
  - There may be uses-cases where domain adaptation is sufficient => identifying scope/objective is important.
- There is a lot to learn from successes in other domains (e.g. vision, language, and robotics) that can be useful in wireless communications
  - Few-shot learning, meta-learning
  - Self-supervised learning techniques for time series data
- Wireless data representation learning and neural architectures matter.
  - A lot of research in the NLP and vision community made advances in these areas first before reaching what we have today.

Standardized real world datasets, challenges, and reproducible research will make the difference to push the field forward!

Thank You!

Questions?

hatem.abouzeid@ucalgary.ca

UNIVERSITY OF CALGARY