Self-supervised Anomaly Detection for Narrowband SETI

Yunfan Zhang

WITH KI HYUN, SEUNGWOO SON, STEVE CROFT AND ANDREW SIEMION

IEEE GlobalSIP 2018

Predictive Anomaly Detection

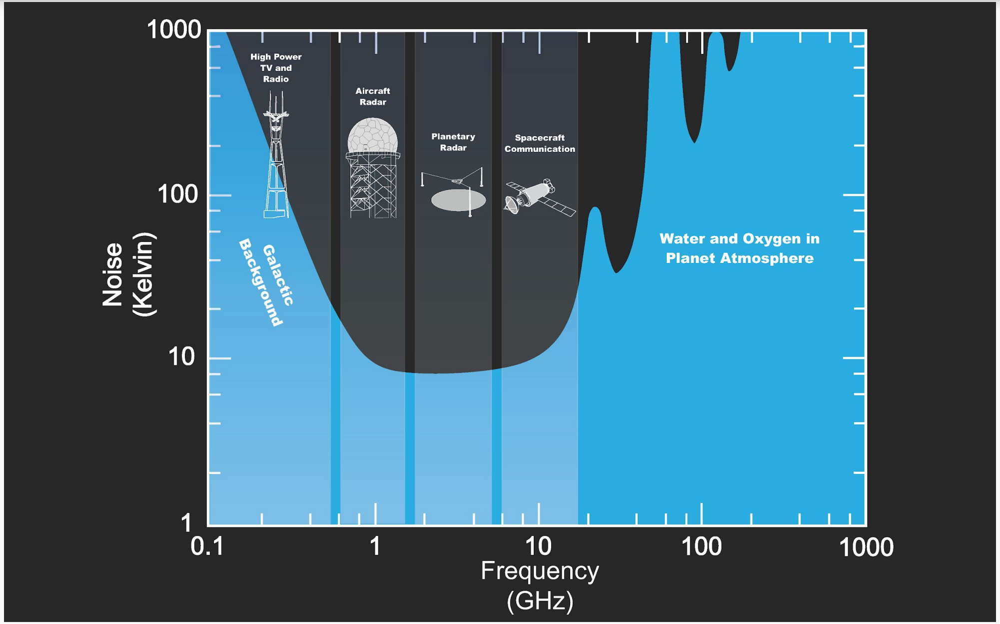
- Spectrogram domain deep learning
- Conv-LSTM GAN
- Applicable to wider range of signals

Outline

1. Context and challenges of narrowband SETI
2. Problem formulation: deep anomaly detection
3. Architecture and design
4. Results and evaluation

Context:
Search for Extra-terrestrial Intelligence

- Technological signals from advanced civilizations.
- Radio band of transparency.
- Main challenges:
  - Unknown signal of interest
  - RFI: Crowded spectrum
  - Large unlabeled dataset
- Algorithm with minimal human supervision



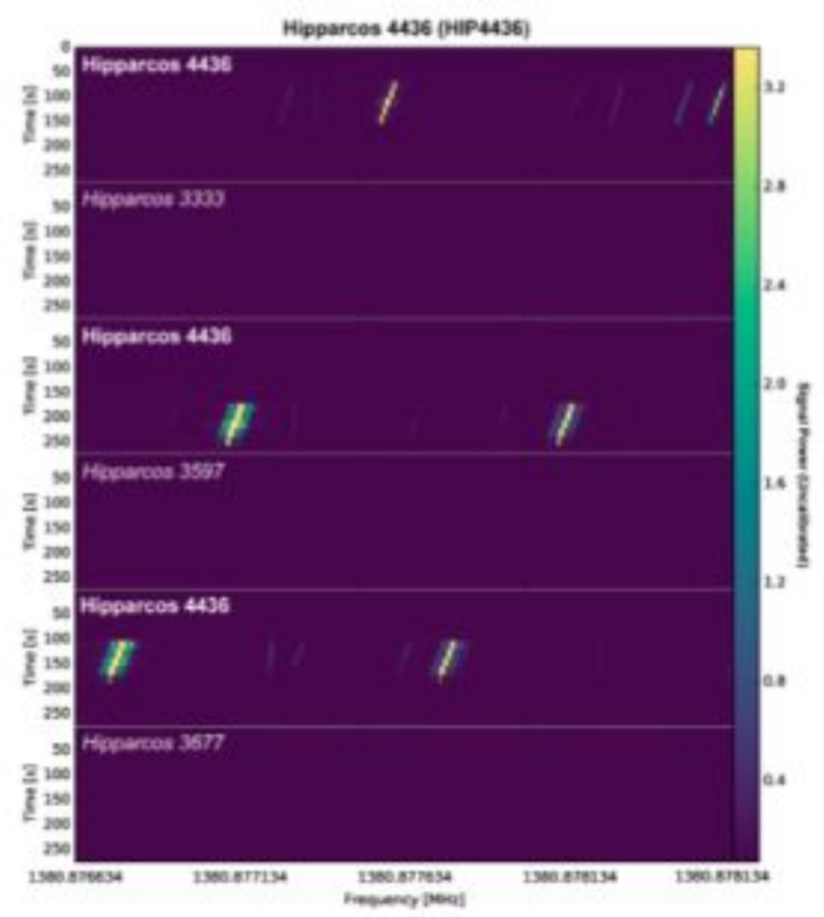
Source: seti.berkeley.edu

Narrowband SETI

- Motivation
  - Energy efficient "attention getter"
  - No natural counterpart

  Algorithmically simple

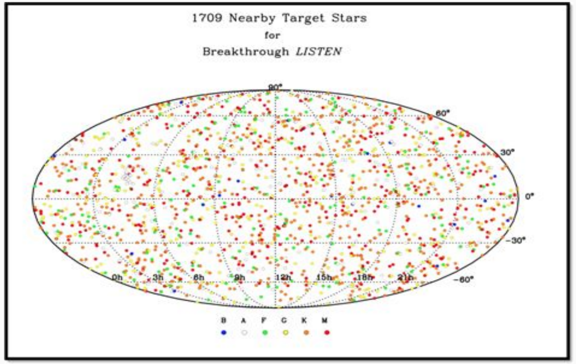
  - Spatial filter for identification
  - Traditional energy detection inflexible

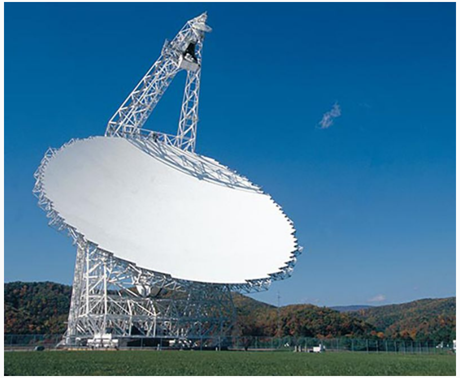


Source: [1]

Breakthrough Listen

- Telescopes: Green Bank Telescope, Parkes Telescope, Meerkat Array
- Mission: 1 million stars, 100 galaxies narrowband search.





- Data rate: 1PB/day IQ, 10 GHz bandwidth
- Form spectrograms of 3 different resolutions in real time, average by factor of 100

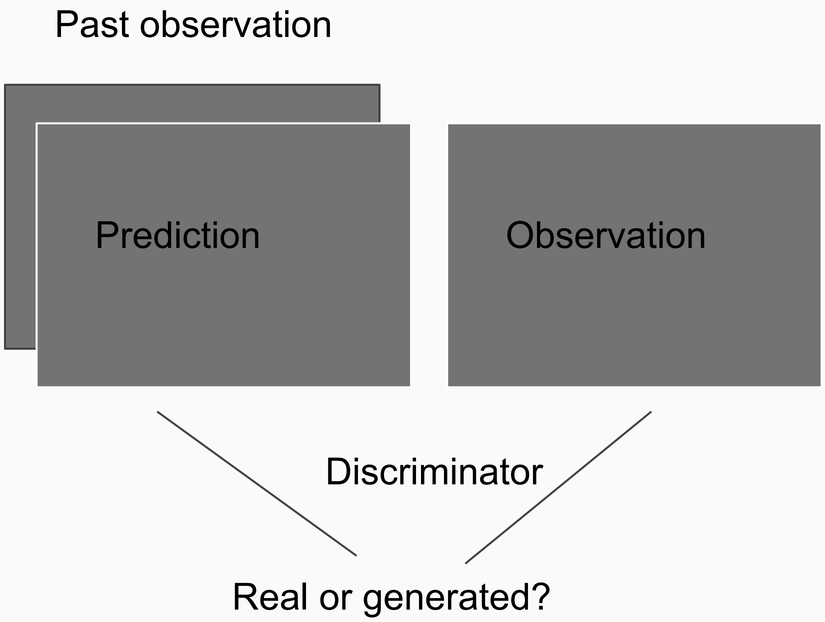
Source: IOP Science

Problem Formulation: Deep Anomaly Detection on Spectrograms

- Detect anomalies by predicting future observations
- Spectral-temporal location of anomaly
- RFI filtering in same framework.
- Time series prediction: RNN and LSTM
- Spatial/frequency dimension: convolution
- Challenge: noise is not predictable
- Solution: introduction discriminator

$$L_g = \log(1\text{-}D(G_{\text{future}}))$$

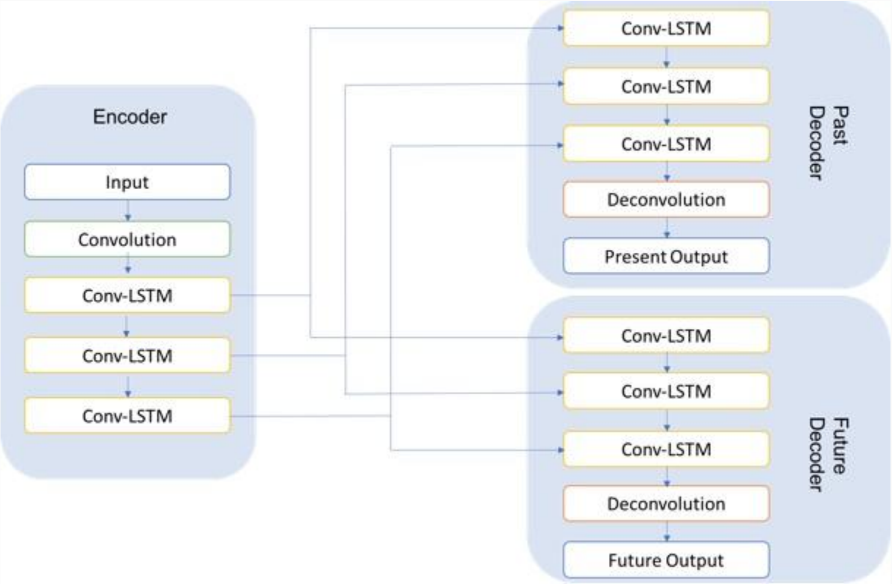
$$L_d = \log(D(G_{\text{future}})) + \log(1\text{-}D(x_{\text{future}})).$$



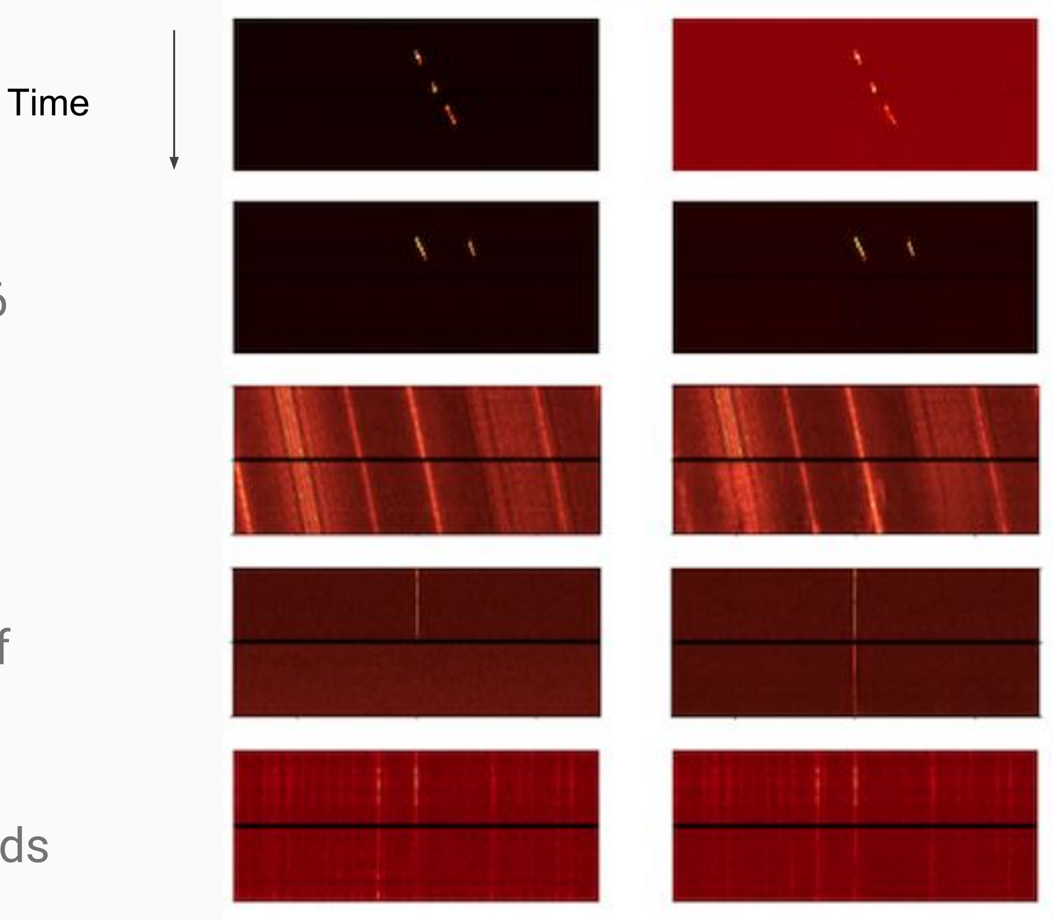
Architecture

- Convolutional LSTM baseline
- Dual decoder
  - Better representation
  - Learn data distribution
- Multiple frames at a time
- Generative Adversarial Loss
  - Regulated training to counter instability

$$L_G = \alpha(L_{\ell 2\text{-future}} + L_{\ell 2\text{-past}}) + \beta L_{\ell 2\text{-feature}} + L_g,$$



Prediction Results



Dataset:

20000 instances of 256 X 16 candidate spectrograms.

Advantages:

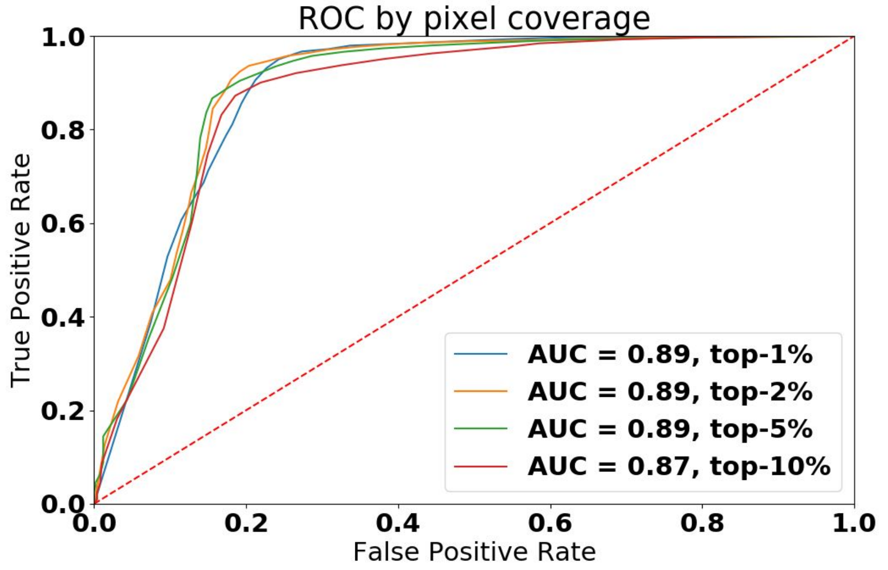
- High fidelity prediction
- Understands discontinuity of signals
- Candidate selection
- Self-supervised learning needs no human labels

Anomaly Detection Evaluation

Pair correspondence with top pixel coverage:

$$\tau \underset{H_0}{\overset{H_1}{\gtrless}} \frac{\|m_1 \& m_2\|}{\|m_1 | m_2\|},$$

False positives due to selection criterion, not prediction model.



Conclusion

- SETI is a challenging effort due to RFI, large data volume, and unknown signal of interest
- Spatial filtering and anomaly detection can both be framed in terms of a generative model.
- We introduce a convolutional LSTM and generative adversarial model to tackle the noisy spectrogram domain.
- Results show promise and possibility of generalizing to other signals.

Thank you!

Contact:

yunfanz@berkeley.edu
seti.berkeley.edu

References:

[1]: J. E. Enriquez, et. al. "The Breakthrough Listen Search for Intelligent Life: 1.1-1.9 GHz Observations of 692 Nearby Stars," ApJ vol. 849, pp. 104, Nov. 2017

Extensions:

- SETI: other signals
- Other RF domains:
  - Anomaly detections (jamming, chirp etc.)