Impact of Localization Errors on Deep Learning Models for Wideband Spectrum Monitoring

Aman Kumar, J. Vikrant, S. Vanka, and S. S. Channappayya
Department of Electrical Engineering
Indian Institute of Technology – Hyderabad, Kandi, Telangana 502 284, India
Email: {ee22btech11006, ee22btech11025}@iith.ac.in; {sundar.vanka, sumohana}@ee.iith.ac.in

***Abstract*—Isolating and classifying of signals of interest in wideband signal traces necessitates distinguishing bursty transmissions from background noise and interference under frequency-selective fading. We find that errors in signal localization that are fundamentally unavoidable in realistic processing pipelines, significantly degrade the classification accuracy of downstream state-of-the-art deep learning-based RF signal classifiers. Our investigation reveals the root cause of this degradation to be a distribution shift between the training and evaluation datasets, particularly within classification windows at the start and end of signal bursts. Importantly, this phenomenon is not specific to the particular localization or classification methods used nor to the use of synthetic datasets; indeed, any practical localization algorithm introduces boundary uncertainty, which leads to a distribution shift between training and inference. These results highlight the need for joint localizer and classifier design for practical real-time spectrum monitoring systems.**

***Index Terms*—Signal classification, Spectrum Monitoring, Edge-AI, Transformer Models, Distribution shift**

I. Introduction

A. Motivation and Related Work

Real-time spectrum monitoring and signal classification requires wideband receivers to increase the rate at which large spectral regions can be covered. In such settings, multiple concurrent transmissions may occur in the presence of background noise. In a practical monitoring system, the time-frequency boundaries of individual packets are *estimated* by a *localizer* to reject unwanted noise and extract individual packets for further downstream processing.

However, state-of-the-art signal classifiers such as AMC-Net [1] and T-PRIME [2] are typically trained and evaluated under the assumption of perfect packet boundaries. In practical deployments where the signal traces are captured over-the-air, the input to these classifiers is packets recovered from the *estimates* rather than actual time-frequency packet boundaries. Several approaches for packet localization have been explored, including energy detection [3], [4], correlation with known preambles [5], [6], and exploiting cyclostationarity [7], and signal area estimation methods [8]–[10] that use edge detection and the Hough transform. Another line of work proposes the use of unsupervised segmentation techniques (e.g., density-based clustering algorithms, such as DBSCAN) on spectrogram images [11].

S. Vanka gratefully acknowledges the support of Qualcomm 6G University Research Funding Program, grant 4300063196.

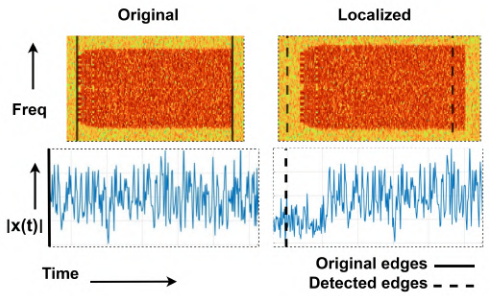


Fig. 1: A spectrogram segment illustrating a packet (red) in the presence of background noise (orange), highlighting the difference between ground-truth (solid black lines) and the boundaries estimated by the localizer (dashed black lines).

However, to the best of our knowledge, little is known about how localizer-induced mismatches between the training and evaluation datasets impact signal classification performance. A key observation motivating this work is that localization errors as in Fig. 1 are *fundamentally unavoidable* due to noise, fading and finite time-frequency resolution. This holds regardless of the data: synthetically generated or captured over the air. Therefore, any localization or object detection algorithm, whether classical or learning-based, introduce boundary uncertainty, leading to this distribution shift. To demonstrate this concretely, we focus on IEEE 802.11 variants as a representative family of bursty wideband signals. However, the observed effects are expected to hold for any wideband bursty transmission subject to localization uncertainty. In this paper, we take the first steps toward investigating this effect. In particular:

- We show that state-of-the-art single-packet protocol classifiers are highly sensitive to time-localization errors, particularly near packet boundaries, leading to significant accuracy degradation.
- We quantify the resulting distribution shift between training and localized test data using the Optimal Transport Dataset Distance (OTDD) [12], showing that localization errors move the test distribution far from the training distribution.

Throughout this work, we use controlled experimental procedures and specific SOTA classifiers to isolate the impact of imperfect localization on downstream classification. These procedures should be viewed as an analytical tool and not as a design prescription.

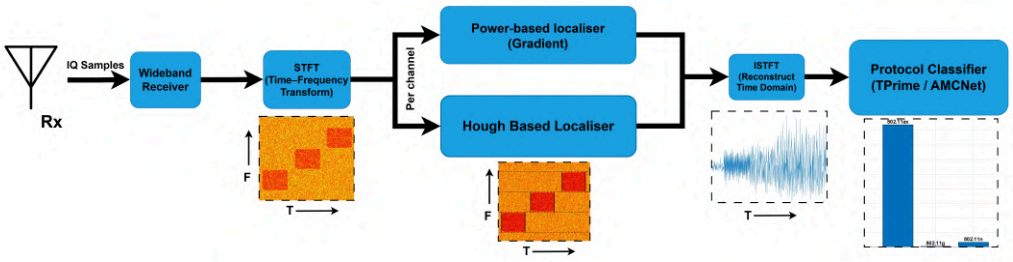


Fig. 2: Wideband classification pipeline showing reception, packet localization, and protocol classification.

II. System Model and Methods for Localization and Classification

A. Wideband Classification Architecture

A practical wideband protocol classification system can be described as a sequence of processing blocks: a wideband receiver, a time–frequency transform, a packet localizer, and a protocol classifier as in Fig. 2. The wideband receiver captures IQ samples over a broad frequency range by simultaneous observation of multiple active channels. The short-time Fourier transform (STFT) is then applied to these samples, converting the received signal into a spectrogram representation that makes localized spectral activity visually apparent in the time–frequency domain. The output of the localizer is a set of time–frequency regions, each ideally corresponding to a single packet. These extracted signals are then provided as input to a protocol classifier, which assigns a label (e.g., Wi-Fi 802.11 variant) to each localized packet. This modular design decouples packet detection from classification and allows the performance of each block to be evaluated separately. This setup provides a concrete framework to evaluate how packet boundary error impacts the classifier stage. Furthermore, for a concrete context for this investigation, we use pre-trained SOTA classifiers for IEEE 802.11 packets [1], [2] and pair them with simple localization (Section II-B) and matching algorithms (Section III-B).

B. Packet Localization Algorithms

We consider two packet localization strategies that operate on spectrograms for each channel: a Hough-transform-based method and a power-gradient-based method. Both methods aim to detect vertical edges in the time–frequency representation, i.e., packet start and end boundaries.

1) Hough-based method: Building on prior work in signal area estimation (SAE) [10], the spectrogram is first enhanced using an edge detector such as the Canny algorithm [13]. The Hough transform then maps edge points to a parameter space, where vertical lines in the spectrogram are detected as peaks. This approach is effective in low-SNR conditions where packet boundaries are faint. It typically achieves good accuracy when edges are well defined. However, its reliance on edges makes it sensitive to frequency selective fading, which can obscure portions of the packet boundary.

2) Power-gradient based method: This approach instead exploits temporal changes in the received signal power. The spectrogram is averaged across frequency bands of 20 MHz, normalized, and smoothed with a Gaussian filter. The packet start and end times are then identified as threshold crossings in the gradient of the smoothed power envelope. Unlike the Hough method, the power-gradient approach does not rely on explicit edge detection and is therefore more resilient to frequency-selective fading. However, it is less robust to noise and may generate extra edges when spurious power fluctuations are present as explored in spectrum sensing and localization contexts [8], [9].

Together, these two methods offer complementary trade-offs: Hough-based localization offers robustness to noise but degrades in multipath, while power-gradient methods remain stable in multipath but are more noise-sensitive.

C. Protocol Classifiers

The localized signals are classified using state-of-the-art deep learning models. We focus on two SOTA architectures:

- AMCNet [1]: a convolutional network designed for automatic modulation classification. Its compact architecture makes it computationally efficient and suitable for real-time spectrum monitoring tasks on edge devices.
- T-PRIME [2]: a transformer-based architecture designed for protocol identification. By leveraging the self-attention mechanism, it captures both local and global temporal dependencies, enabling strong performance across diverse SNR conditions.

III. Protocol Classification with Imperfect Localization

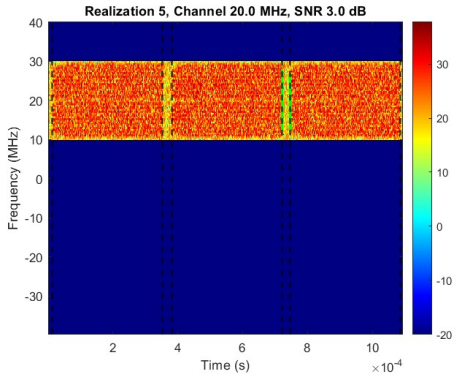
A. Experimental Setup and Data Generation

To study the impact of imperfect localization on protocol classification, we construct a synthetic wideband dataset and evaluate the classifier performance on packets detected by the localizer. Each realization consists of a multichannel time-domain trace spanning an 80 MHz band, formed by three 20 MHz sub-bands centered at -20, 0, and $+20$ MHz. Each sub-band contains a random number of packets with randomized inter-packet intervals, ensuring no temporal overlap. Thus the total inter-frame spacing varies in the range:

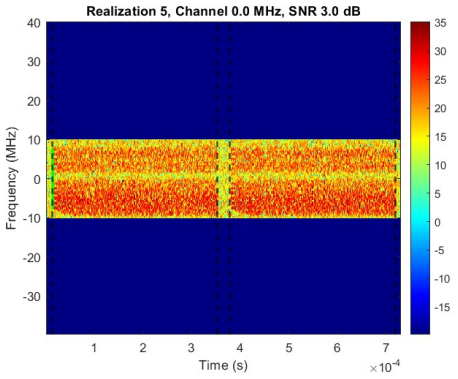
$$T_{\text{gap}} = 16\ \mu\text{s} + 9 \times U[0, 10]\ \mu\text{s} \sim U(16\mu s, 106\mu s).$$

where, 9 is the IEEE 802.11 slot time [14].

Packet waveforms are generated using IEEE 802.11 PHY configurations. The resulting dataset captures realistic variability in packet placement and timing, enabling controlled evaluation of localization-induced classification errors.



(a) Ch. 20 MHz: 3 packets with 16 µs gap. Occlusion prevents Hough from detecting lines.



(b) Ch. 0 MHz: Multipath knocks off frequency components, degrading Hough detection.

Fig. 3: Hough line detection (Realization 5, 3 dB SNR) showing degradation from occlusion, multipath, and spectral loss.

We then apply two channel conditions to each packet: AWGN and TGn multipath (Model D) [15], with SNR values selected from $\{0, 5, 10, 15, 20, 25, 30\}$ dB and applied to each 20 MHz packet, resulting in different SNRs across the trace.

B. Detected Edge Matching and localization errors

Consider the ground-truth vertical edges

$$\mathcal{G} = \{g_1, g_2, \ldots, g_n\}, \quad g_1 < g_2 < \cdots < g_n, \tag{1}$$

and the localizer outputs m inferred edges

$$\mathcal{D} = \{d_1, d_2, \ldots, d_m\}, \quad d_1 < d_2 < \cdots < d_m, \tag{2}$$

The localizer operates without using start or end labels at inference time, all the detected edges are treated as unlabelled vertical edges. To evaluate the localization performance, we perform a greedy one-to-one matching between detected and ground-truth edges.

Greedy matching procedure: Starting from the first elements of both ordered sets $\mathcal{G}$ and $\mathcal{D}$, the algorithm proceeds as:

1) Initialise $i = j = 1$.
2) While $i \leq n$ and $j \leq m$:
   - Compute the absolute difference $\Delta = |g_i - d_j|$.
   - If $\Delta \leq \tau$, pair (g_i, d_j) as a match, else, do not pair.
   - Increment $i \leftarrow i + 1$ and $j \leftarrow j + 1$.

Thus, only the first $k = \min(m, n)$ pairs satisfying $|g_i - d_j| \leq \tau$ are considered valid matches, ensuring that the matches are both greedy and order preserving.

The greedy procedure assumes edges are sufficiently sparse, which is typically enforced by the 802.11 channel access mechanism. When this does not hold, advancing both pointers on a mismatch can cascade, pairing a ground-truth edge with a distant detected edge. This controlled synthetic setting is designed to avoid such cases; handling insertions and deletions via optimal assignment is left for future work.

We consider the following localization error that can occur:

Missing or extra edges: These occur when the number of detected edges does not match the number of ground-truth edges. Unmatched ground-truth edges correspond to false negatives (FN), and unmatched detections correspond to false positives (FP). The matched pairs are counted as true positive (TP):

$$\text{TP} = k, \text{FP} = \max(0, m - n), \text{FN} = \max(0, n - m), \tag{3}$$

The standard detection metrics are then computed as:

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}}, \qquad \text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}}. \tag{4}$$

Even when edges are correctly matched, detection may occur slightly before or after the true boundary, introducing boundary offset errors that cause a distributional shift in the classifier input.

IV. Classification under Imperfect Localization

A. Localizer Performance

We evaluate the performance of both the localization algorithms by summarizing Precision/Recall tables across varying SNR and channel conditions.

1) Hough Transform-based method: The Hough-based localizer maintains a high TP rate and low FP rate across all SNRs in both AWGN and Rayleigh fading scenarios (Tables I and II). However, its recall degrades under multipath, as distortion in packet boundaries in the spectrogram weakens the vertical edge structure that the Hough transform relies on. Figure 3(b) illustrates this effect at 3 dB SNR under frequency-selective fading, where multipath causes selective attenuation of certain frequency components, leading to missed detections of packet boundaries. An advantage of the Hough transform is that it can detect a line of any orientation without prior knowledge of channel boundaries. In this work, we apply it *per channel* to improve detection reliability, allowing better separation of signals that are close in time but lie in different bands, reducing the chances of missed edges or merging multiple sub-band packets into a single packet.

SNR (dB)	Ch −20		Ch 0		Ch 20	
	Precision	Recall	Precision	Recall	Precision	Recall
-3	0.925	0.870	0.916	0.838	0.928	0.798
0	0.949	0.964	0.946	0.919	0.957	0.878
3	0.995	0.974	0.996	0.914	0.999	0.848
6	1.000	0.977	1.000	0.939	1.000	0.846
9	1.000	0.984	1.000	0.945	1.000	0.860
12	1.000	0.983	1.000	0.945	1.000	0.842

TABLE I: Shape-based Method (no multipath): Precision/Recall vs. SNR.

SNR (dB)	Ch −20		Ch 0		Ch 20	
	Precision	Recall	Precision	Recall	Precision	Recall
-3	0.968	0.539	0.968	0.521	0.973	0.504
0	0.980	0.725	0.984	0.703	0.983	0.670
3	0.999	0.803	0.999	0.798	0.998	0.747
6	0.999	0.858	1.000	0.850	1.000	0.809
9	1.000	0.900	1.000	0.879	1.000	0.912
12	1.000	0.935	1.000	0.911	1.000	0.893

TABLE II: Shape-based Method (multipath): Precision/Recall vs. SNR.

SNR (dB)	Ch −20		Ch 0		Ch 20	
	Precision	Recall	Precision	Recall	Precision	Recall
-3	1.000	0.856	0.998	0.855	0.999	0.859
0	0.996	1.000	0.996	1.000	0.998	1.000
3	0.998	1.000	0.997	1.000	0.996	1.000
6	0.995	1.000	0.996	1.000	0.998	1.000
9	0.996	1.000	0.998	1.000	0.993	1.000
12	0.997	1.000	0.996	1.000	0.998	1.000

TABLE III: Power-based Method (multipath): Precision/Recall vs. SNR.

2) Power-gradient based method: In contrast, the power-gradient detector achieves higher recall (Table III) even with multipath, it is about 0.85 at -3 dB and becomes 1 above 0 dB for all channels. This indicates that it detects most true edges, but occasionally produces extra edges, slightly lowering precision.

Overall, the Hough-based localizer is sensitive to multipath effects, leading to significant recall degradation even at high SNRs whereas the power-based localizer is more robust under multipath and low-SNR conditions, maintaining higher recall. Since realistic traces include both faded and unfaded packets, we use the power-based localizer in subsequent analysis. These observations highlight that some level of localization error is unavoidable regardless of the method used.

B. Classifier Performance under Imperfect Localization

To evaluate the impact of imperfect boundaries on classification, we use packets detected by the power-based localizer and fed to the two pre-trained classifiers: T-PRIME and AMCNet. Each packet is segmented according to the required window size, and final decisions are obtained by averaging segment-level predictions.

The confusion matrices in Tables IVA and VA show that, under multipath conditions, both classifiers exhibit noticeable misclassification between the IEEE 802.11ax and IEEE 802.11n protocols, while maintaining high confidence when classifying IEEE 802.11g. This suggests that certain waveform types are more sensitive to boundary errors. This behavior can be understood in terms of localization errors. While precision and recall measure detection accuracy, they do not capture temporal misalignment. In our setting, most detected edges are slightly shifted from the true boundaries, causing the initial and final segments of a packet contain a mixture of noise and signal samples, which differ from the training distribution. In addition, the lower precision of the power-based localizer leads to extra edge detections, resulting in noise-only packets being passed to the classifier. To study this effect, we analyze classifier responses to noisy segments under multipath conditions and identify two cases.

	Prediction		
Actual	**ax**	**g**	**n**
ax	0.772	0.061	0.168
g	0.000	0.980	0.020
n	0.004	0.083	0.913

(A) TGn Multipath

	Prediction		
Actual	**ax**	**g**	**n**
ax	0.999	0.000	0.001
g	0.000	1.000	0.000
n	0.000	0.000	1.000

(B) No Multipath

TABLE IV: Confusion matrices of T-PRIME: Overall performance for TGn and No Multipath.

	Prediction		
Actual	**ax**	**g**	**n**
ax	0.897	0.011	0.092
g	0.002	0.953	0.045
n	0.001	0.018	0.981

(A) TGn Multipath

	Prediction		
Actual	**ax**	**g**	**n**
ax	1.000	0.000	0.000
g	0.000	0.994	0.006
n	0.000	0.000	1.000

(B) No Multipath

TABLE V: Confusion matrices of AMCNet: Overall performance for TGn and No Multipath.

1) Mixed Noise-Signal Segments: These segments arise due to boundary offset errors and contain noise and signal samples. In this case, both classifiers tend to confuse ax with n, while maintaining higher confidence for g, as shown in Tables VIA and VIIA, indicating that not all legacy waveforms are equally impacted by localization imperfections.

2) Noise-only Segment: Segments consisting purely of noise samples arise due to extra edge detections (precision < 1), as shown in Tables VIB and VIIB. These reveal a systematic decision bias, where the classifier defaults to its baseline behavior, ($ax > n > g$). It is also observed that the relatively simpler AMCNet in Table VIIB is more prone to this default decision issue than T-PRIME in Table VIB. This behavior is not due to class imbalance, but instead stems from the structural complexity of the ax class.

To better understand this degradation at the dataset level, we next quantify how localization alters the data distribution seen by the classifier.

C. Effect of localization on dataset distribution (OTDD)

Residual noise due to boundary errors can change the statistical structure of the features fed to the classifier. To quantify this effect, we measure the induced data distribution

Actual	Prediction		
	ax	g	n
ax	0.184	0.191	0.625
g	0.012	0.897	0.092
n	0.044	0.544	0.412

(A) Noise + Signal

Actual	Prediction		
	ax	g	n
ax	0.528	0.160	0.312
g	0.528	0.160	0.312
n	0.528	0.160	0.312

(B) Noise Only

TABLE VI: Confusion matrices T-PRIME: 1st segment performance under mixed and pure noise signals.

Actual	Prediction		
	ax	g	n
ax	0.309	0.024	0.667
g	0.014	0.866	0.119
n	0.023	0.273	0.704

(A) Noise + Signal

Actual	Prediction		
	ax	g	n
ax	0.735	0.046	0.219
g	0.735	0.046	0.219
n	0.735	0.046	0.219

(B) Noise Only

TABLE VII: Confusion matrices of AMCNet: 1st segment performance under mixed and pure noise signals.

shift using the Optimal Transport Dataset Distance (OTDD) [12].

The analysis was performed on the same 1000 realization corpus used for evaluating the localizer performance, resulting in about 7000 packets. For each packet, the first segment is extracted and flattened by concatenating the I and Q channels, consistent with the classifier input. Ground-truth packet boundaries define the training set, while packets detected by the localizer form the test set (≈ 6500 packets).

a) OTDD configuration: We computed the OTDD using the debiased Sinkhorn formulation of the 2-Wasserstein distance with a small entropy regularisation term. The cost between samples is based on the squared Euclidean distance in the high dimensional feature space. No label information is used, so the OTDD reflects purely geometric differences between the datasets.

b) Results and Interpretation: We compare three dataset pairs to quantify the distributional shift introduced by localization:

- Train set vs. the power-based localizer output gave us an OTDD value of ≈ 1600.
- To establish a reference for a large shifted distribution, a synthetic dataset of *AWGN* bursts was generated, which after localization, gives an OTDD of ≈ 1550 compared to the training set.
- To estimate the internal variation of our training set, we randomly split the 7000 training packets into two non-overlapping subsets of equal size resulting in a much smaller OTDD of approximately 300.

The OTDD between the training data and the localized packets ($\approx$ 1600) is nearly identical to that between the training data and pure noise ($\approx$ 1550). This shows that even small localization offsets introduce a significant second-order (covariance) effect in the feature space. Rather than shifting the global mean, imperfect segmentation alters the spread and orientation of the data in the high-dimensional manifold, increasing the transport cost. In contrast, the much smaller OTDD between two disjoint subsets of the training set ($\approx$ 300) reflects only internal variation. Together, these observations indicate that localization-induced covariance distortions in the test distribution are a primary reason for degraded classifier performance.

V. Conclusion and Future Work

This work highlights the importance of building models that can handle noisy wireless signals and uncertain packet boundaries. Using power-gradient and Hough transform localization, we observe reliable packet detection and Wi-Fi classification even in low SNR and multipath conditions. However, OTDD analysis showed that poor generalization results from distributional shifts introduced by localization errors, with some waveforms affected more than others. Future work should focus on training classifiers directly on realistic localization outputs so they generalize across localizers, extending the approach to other wideband signals.

References

[1] J. Zhang *et al.*, "Amc-net: An effective network for automatic modulation classification," in *ICASSP 2023-2023 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2023, pp. 1–5.

[2] M. Belgiovine *et al.*, "T-prime: Transformer-based protocol identification for machine-learning at the edge," in *IEEE INFOCOM 2024-IEEE Conference on Computer Communications*. IEEE, 2024, pp. 1031–1040.

[3] S. Atapattu *et al.*, *Energy Detection for Spectrum sensing in Cognitive Radio*. Springer, 2014, vol. 6.

[4] M. A. Abdulsattar *et al.*, "Energy detection technique for spectrum sensing in cognitive radio: a survey," *International Journal of Computer Networks & Communications*, vol. 4, no. 5, p. 223, 2012.

[5] S. Banerji *et al.*, "On IEEE 802.11: Wireless LAN Technology," *arXiv preprint arXiv:1307.2661*, 2013.

[6] K. Pahlavan *et al.*, "Evolution and impact of Wi-Fi technology and applications: A historical perspective," *International Journal of Wireless Information Networks*, vol. 28, pp. 3–19, 2021.

[7] Y. Zeng *et al.*, "Spectrum-sensing algorithms for cognitive radio based on statistical covariances," *IEEE Transactions on Vehicular Technology*, vol. 58, no. 4, pp. 1804–1815, 2008.

[8] M. M. Alammar *et al.*, "Evaluation of the impact of thresholding and frequency/time resolution on signal area estimation methods," in *2021 IEEE 93rd Vehicular Technology Conference (VTC2021-Spring)*. IEEE, 2021, pp. 1–7.

[9] M. M. Alammar, "Enhanced signal area estimation based on edge detection and flood fill," *IEEe Access*, vol. 10, pp. 47 179–47 194, 2022.

[10] M. M. Alammar *et al.*, "Automatic extraction of signal areas from radio spectrograms based on the hough transform," in *2022 IEEE 23rd International Symposium on a World of Wireless, Mobile and Multimedia Networks (WoWMoM)*. IEEE, 2022, pp. 204–213.

[11] M. López-Benítez *et al.*, "Accurate automatic extraction of signal components from noisy radio spectrograms," *IEEE Transactions on Cognitive Communications and Networking*, vol. 8, no. 4, pp. 1604–1617, 2022.

[12] D. Alvarez-Melis *et al.*, "Geometric dataset distances via optimal transport," *Advances in Neural Information Processing Systems*, vol. 33, pp. 21 428–21 439, 2020.

[13] J. Canny, "A computational approach to edge detection," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, no. 6, pp. 679–698, 1986.

[14] "Ieee standard for information technology–telecommunications and information exchange between systems local and metropolitan area networks–specific requirements part 11: Wireless lan medium access control (mac) and physical layer (phy) specifications amendment 1: Enhancements for high-efficiency wlan," *IEEE Std 802.11ax-2021 (Amendment to IEEE Std 802.11-2020)*, pp. 1–767, 2021.

[15] V. Erceg *et al.*, "TGn Channel Models," IEEE 802.11 Wireless LANs, Tech. Rep. IEEE 802.11-03/940r4, May 2004, [Online]. Available: https://mentor.ieee.org/802.11/dcn/09/11-09-0308-00-00ac-tgac-channel-model-addendum-document.doc.