

# EVALUATING LEARNING-BASED AND CLASSICAL SUPER-RESOLUTION

## FOR ACOUSTIC FMCW SENSING ON SMARTPHONES

Aryan Mahindra

University of Washington · CSE 493G1: Deep Learning · Winter 2026

### Motivation

- Smartphones can sense nearby surfaces using near-ultrasonic FMCW chirps (16–20 kHz), adapting radar-like ideas to sound
- With 4 kHz bandwidth, the nominal Rayleigh range resolution is about **4.29 cm**
- Can learning-based or model-based methods recover finer *apparent* structure than this bandwidth-limited baseline?

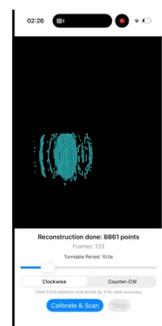
$$\Delta r = \frac{c}{2BW} \approx 4.29 \text{ cm}$$

- **Input:** 141 complex samples over 0–50 cm
- **Goal:** recover a 1128-sample profile (8× finer grid)
- **Constraint:** exact data consistency with the measured samples

### Hardware



Object on turntable



3D point cloud, 8,861 pts

### Related Work

**FingerIO** (CHI 2016): acoustic finger tracking, mm-scale localization.

**MilliSonic** (CHI 2019): FMCW acoustic tracking, sub-mm localization.

**NOMP** (TSP 2016): model-based spectral super-resolution with CFAR stopping.

### Approach 1: NullspaceNet (DL)

*Complex 1D CNN, about 1.04M parameters*

1. Baseline:  $\mathbf{x}_0 = A^\dagger \mathbf{y}$
2. CNN correction:  $\mathbf{r}_{\text{raw}} = \text{CNN}(\mathbf{x}_0)$
3. Nullspace projection:

$$\mathbf{r}_{\text{null}} = \mathbf{r}_{\text{raw}} - A^\dagger A \mathbf{r}_{\text{raw}}$$

4. Output:

$$\mathbf{x} = \mathbf{x}_0 + \mathbf{r}_{\text{null}}$$

By construction,  $A\mathbf{x} = \mathbf{y}$ . The network can only modify nullspace components.

**Loss:**

$$L = w_1 \|x - x_{\text{gt}}\|_1 + w_p L_{\text{peak}} + w_E \|\mathbf{r}_{\text{null}}\|^2$$

### Approach 2: NOMP (Classical)

*Zero learned parameters*

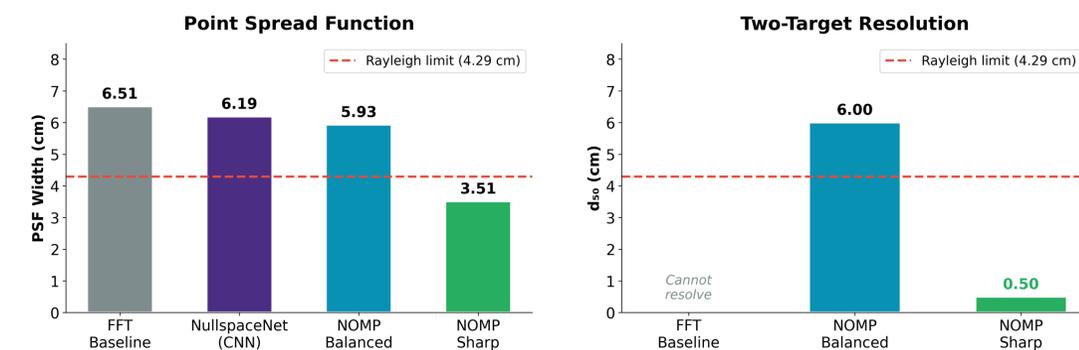
Key idea: estimate sparse reflectors at **continuous** positions, rather than regressing a dense profile.

1. Search the grid for the strongest reflection (GLRT)
2. Refine its location with Newton updates
3. Subtract the reflector from the residual
4. Repeat with cyclic refinement
5. Stop using a CFAR threshold

Measured samples remain exactly matched, while the unobserved high-resolution structure is inferred from a sparse reflector model.

### Results

120 synthetic scenes, SNR = 20 dB, 1–15 reflectors per scene



Method	PSF (cm)	FP Rate	MD Rate	$d_{50}$ (cm)
FFT Baseline	6.51	0.736	0.194	N/A
NullspaceNet (CNN)	6.19	~0.736	~0.18	~4.0
<b>NOMP Sharp</b>	<b>3.51</b>	<b>0.634</b>	0.249	<b>0.50</b>
NOMP Balanced	5.93	0.694	<b>0.172</b>	6.00

On sparse synthetic scenes, NOMP Sharp reaches  $d_{50} = 0.50 \text{ cm}$ , resolving separations below the nominal Rayleigh limit under the assumed model

NOMP Sharp trades higher missed detection for stronger separation and fewer false positives. NOMP Balanced favors lower missed detection at the cost of coarser separation.

### Why Dense Regression Struggled

- **Output mismatch:** 1128 dense samples for a few reflectors creates many valid solutions
- **Loss bias:**  $L_1$  favored smoother, lower-risk outputs
- **NOMP advantage:** estimates count and location directly

### Conclusion & Future Work

**In this experiment, formulation mattered more than scaling the CNN**

The CNN preserved data consistency but did not improve separation as much as NOMP. Real-world examples show qualitatively improved reconstruction.

**Next:** neural-initialized NOMP, bandwidth extension via harmonics, set-based prediction.