Super-Resolution Time-of-Arrival Estimation using Neural Networks

Yao-Shan Hsiao*, Mingyu Yang*, and Hun-Seok Kim
EECS, University of Michigan, Ann Arbor, MI, USA

***Abstract*—This paper presents a learning-based algorithm that estimates the time of arrival (ToA) of radio frequency (RF) signals from channel frequency response (CFR) measurements for wireless localization applications. A generator neural network is proposed to enhance the effective bandwidth of the narrowband CFR measurement and to produce a high-resolution estimation of channel impulse response (CIR). In addition, two regressor neural networks are introduced to perform a two-step coarse-fine ToA estimation based on the enhanced CIR. For simulated channels, the proposed method achieves 9% – 58% improved root mean squared error (RMSE) for distance ranging and up to 22% improved false detection rate compared with conventional super-resolution algorithms. For real-world measured channels, the proposed method exhibits an improvement of 1.3m in distance error at 90 percentile.**

***Index Terms*—time-of-arrival (ToA) estimation, super-resolution, neural networks (NN), deep learning**

I. Introduction

The goal of time of arrival (ToA) estimation for RF based localization is to find the direct (shortest) propagation delay from the transmitter to the receiver, which is proportional to the distance. The theoretical accuracy of the ToA estimation is fundamentally limited by the RF signal bandwidth whereas the bandwidth of practical systems is severely limited due to the interference to other devices, transceiver complexity, and power consumption of the system [1] [2]. When the bandwidth is restricted to tens of MHz as specified in the ISM band regulation, Nyquist sampled channel impulse response (CIR) only provides a very coarse (a few meters) resolution.

A multi-path channel is typically modeled as

$$h(t) = \sum_{l=0}^{L-1} c_l \delta(t - \tau_l) \tag{1}$$

where L is the total number of multi-paths, c_l and τ_l are the complex attenuation and propagation delay of the lth path. Multi-paths are indexed in the ascending order of the propagation delay of each path. Thus, τ_0 is the ToA of the shortest path that reveals the distance between the transmitter and the receiver. In this paper, we focus on estimating τ_0 from the noisy discrete frequency-selective (i.e., multi-path rich) channel frequency response (CFR) $H[n]$, which can be expressed as

$$H[n] = \sum_{l=0}^{L-1} c_l e^{-j2\pi n \frac{B}{N} \tau_l} + w[n], \quad n = 0, \ldots, N-1 \tag{2}$$

This work was supported by the NIST Public Safety Innovation Accelerator Program (PSIAP) Grant 70NANB17H163.

* Equally contributed authors

where N is the total number of samples (subcarriers) in the frequency domain, B denotes the bandwidth and w denotes the noise. $H[n]$ can be measured by orthogonal frequency-division multiplexing (OFDM)-based channel estimation [1] [2]. The simplest way to estimate ToA from $H[n]$ is to compute the corresponding CIR $h[n]$ obtained by inverse Discrete Fourier Transform (IDFT). However, due to the limited bandwidth B, the time resolution of $h[n]$ is essentially limited to $1/B$. That is, when B is small, multi-paths pulse energy leaks to other samples and it becomes difficult to isolate the time of the first path pulse (τ_0) apart. For IEEE 802.11a/g/n WiFi standard, the OFDM operation bandwidth is 20/40MHz and this corresponds to a resolution of $15/7.5$m in distance, which is too coarse for many practical localization systems.

Inspired by the recent remarkable success of deep learning, we propose a two-stage, deep-learning-based algorithm that involves multiple neural networks to estimate the ToA in a finer resolution. First, we train a generator network that 1) extends the effective bandwidth and generates a CIR with higher resolution, and 2) performs de-noising that distinguishes the actual impulses in CIR from the noise. In the next stage, we apply a coarse-fine cascade estimation process, where the generated high-resolution CIR is first fed into a regressor network to get a coarse estimation of the ToA. Then, we train another regressor network that only focuses on the region-of-interest around the coarse ToA to estimate ToA in a finer manner. In the experimental section, we show that the proposed method achieves significant gain from traditional methods in both simulated and real-world measured channels.

II. Background

By exchanging the role of time and frequency in (2), the measured CFR becomes a sum of harmonic signals while the arrival time of paths becomes the corresponding line spectrum. Thus, the ToA estimation can be formulated as a well-known line spectra estimation problem with a goal to estimate the first line in the spectrum (i.e., τ_0). In literature, multiple signal classification (MUSIC) [3] and signal parameters via rotational invariance techniques (ESPRIT) [4] are the two most popular super-resolution techniques and they are often referred as subspace methods. Their applications to ToA estimation have been studied in [5] and [6]. For these approaches, one main issue is that they rely on accurate correlation matrix estimations to maintain high accuracy, which usually requires multiple snapshots of the CIR / CFR measurements. If only one snapshot is available, they need to reduce the efficient

bandwidth to estimate the correlation matrix. Besides, they assume the number of paths as prior information, which is often unavailable in practical systems.

Optimization methods that exploit signal sparsity have been recently proposed for line spectral estimation [7] [8]. One main problem for these methods is that they assume a certain sparsity condition for the signal. Thus the performance severely degrades when the sparsity condition is not met in multipath-rich conditions. Moreover, their complexity can be excessively high as the optimization requires solving a semidefinite program or a LASSO problem that involves a large discrete Fourier transform matrix.

A data-driven method for ToA estimation was originally proposed in [9], where authors focus on finding the leading edge of the magnitude of CIR. This approach estimates the ToA by finding the time index of CIR that best fits the leading edge patterns in the dataset with least square as the similarity metric. In [1] and [2], the CIR leading edge finding problem is solved using a shallow fully connected neural network trained with a simulated CIR dataset. These prior works demonstrate improved ToA accuracy compared to non-neural network based approaches although the gain is limited due to the simple structure of the neural network. More recently, a convolutional neural network (CNN) based approach was demonstrated to solve line-spectra estimation problems [10]. However, its main focus is to analyze line spectras that are relatively far apart from each other in time domain. That assumption does not necessarily hold for ToA-based localization in realistic multipath channels.

III. Proposed Method

A. Overview

Treating a deep neural network as a blackbox and directly training it to produce a desired output (in our case, ToA) with a large dataset often results in poor performance [10]. To outperform conventional (non-neural network) algorithms, it is critical to impose a proper network structure, loss function, and training methodology for deep neural network training. Inspired by recent neural-network-based signal processing approaches [11] [12], we split the task of ToA estimation into two stages: CIR Enhancement Stage and ToA Estimation/Regression Stage. The proposed two-step process is shown in Fig. 1.

1) CIR Enhancement Stage: The goal of this stage is to generate a de-noised CIR with a resolution higher than Nyquist sampling rate. Let $\mathbf{H} = \{H[0], \ldots, H[N-1]\}$ denote the observed CFR, $\mathbf{D}$ denotes the $NR \times NR$ DFT matrix, R denotes the up-sampling rate and $\mathbf{h}_{high} \in \mathbb{C}^{NR}$ denotes the high-resolution CIR. Then, the observation model can be written as:

$$\mathbf{H} = \tilde{\mathbf{D}}\mathbf{h}_{high} + \mathbf{w} \tag{3}$$

where $\tilde{\mathbf{D}} \in \mathbb{C}^{N \times NR}$ only contains the first $N/2$ and last $N/2$ rows of $\mathbf{D}$ while $\mathbf{w}$ represents the noise. Since there are less observations than unknowns, this problem is under-determined and often referred as an inverse problem. Motivated by recent deep-learning-based solutions on inverse problems such as audio/image super-resolution [12]–[14], we train a generator neural network that 'synthesizes' details added to the low-resolution (interpolated to a higher sampling rate) CIR to generate $\mathbf{h}_{high}$. Denoting the generator network by G, the CIR Enhancement Stage operation can be expressed as:

$$\hat{\mathbf{h}}_{high} = G(\tilde{\mathbf{D}}^{\dagger}\mathbf{H}) \tag{4}$$

where $\tilde{\mathbf{D}}^{\dagger}$ is the pseudo-inverse of $\tilde{\mathbf{D}}$ and $\hat{\mathbf{h}}_{high}$ is the estimated high-resolution CIR. Here, $\tilde{\mathbf{D}}^{\dagger}\mathbf{H}$ is essentially an interpolated CIR that can be calculated by zero-padding $\mathbf{H}$ and applying IDFT. In the experimental section, we show that this CIR enhancement method significantly outperforms simple interpolation.

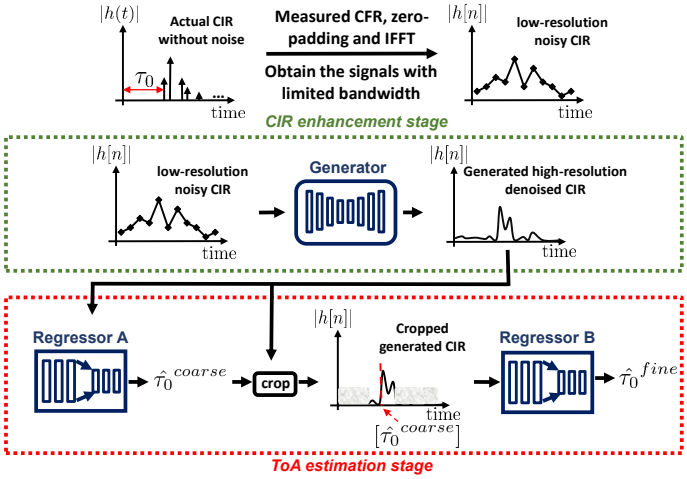

Fig. 1. The proposed two-stage network structure

2) ToA Estimation Stage: In this stage, we treat ToA estimation as a regression problem. Motivated by the cascade estimation method adopted in facial point localization [15] and pose estimation [11], we propose a coarse-fine estimation process where the generated high-resolution CIR is first fed into a CNN-based regressor network (Regressor A) to produce a coarse estimation. Then, we train another regressor network (Regressor B) to learn the displacement of the coarse estimation to the true ToA. Different from Regressor A, Regressor B only focuses on the relevant part of the generated high-resolution CIR; it is applied to the cropped window around the coarse estimation (input is significantly shorter than the full CIR). This approach reduces the distraction to Regressor B caused by noise and multipaths arriving later. In the experimental section, we show that this cascade structure does improve the final ToA estimation accuracy over single regressor (Regressor A only).

Let R_A and R_B denote Regressor A and Regressor B. Then the coarse estimation can be expressed as

$$\hat{\tau_0}^{coarse} = R_A(\hat{\mathbf{h}}_{high}) \tag{5}$$

However, the coarse estimation may not correspond exactly to an integer sample index. Thus, we first find the closest sample index around the coarse estimation (denoted as $[\hat{\tau_0}^{coarse}]$)

and then crop the CIR into a shorter version around that sample point. When the coarse ToA is at the very beginning or the end, we perform cyclic shift to produce the cropped CIR with the same length. Denoting the cropping process by $f_c(.)$, we have

$$\hat{\tau_0}^{fine} = [\hat{\tau_0}^{coarse}] + R_B(f_c(\hat{\mathbf{h}}_{high}, \hat{\tau_0}^{coarse})) \quad (6)$$

B. Neural Network Structure

For the generator, we adopt a Unet-based structure that was originally proposed in [13] for real-valued audio super-resolution problems, which consists of 4 pairs of down-sample and up-sample blocks. For our complex-valued input, we treat the real part and the imaginary part as two separate channels and concatenate them. Each down-sample block contains a 1-D convolutional layer with stride 2 and size-9 (in the first 2 blocks) or size-5 (in the last 2 blocks) kernels that is followed by a leaky ReLU activation. The number of feature maps increases in down-sampling and decreases in up-sampling. Each up-sample block contains a convolutional layer with stride 1, leaky ReLU and a shuffle layer which is used to concatenate the corresponding down-sample input. (Details can be found in [13].)

For both Regressor A and Regressor B, we use a CNN-based estimator that includes 3 convolutional layers followed by 3 fully connected layers. Each 1-D convolutional layer contains 128 channels with small 3×1 kernels, followed by a batch normalization, ReLU and a maxpooling layer that downsamples the input by 2. The first two fully connected layers have 1024 hidden units while the last fully connected layer has an output size of 1 that represents the ToA.

C. Training strategy

The training data for the generator contains pairs of the simulated noisy low-resolution CIR and the noiseless high-resolution CIR that are generated from the channel model described in Section IV. To train the generator, we combine the time-domain amplitude loss L_A and the frequency-domain pointwise loss L_F as shown in (Eq. 7) where both loss are calculated by the L2 norm of the difference between the generated and ground-truth high-resolution CIR. In (Eq. 7), $G(.)$ is the generator A, x is the low resolution interpolated CIR, F is the DFT matrix, $G(x)$ is the generated high-resolution CIR, and y is the ground-truth high-resolution CIR. Here, we observed adding the amplitude error of complex samples in time-domain can lead to a better result than only using the frequency-domain error. An output example of the generator is shown in Fig. 2 to illustrate the de-noising effect and added details in the generated high-resolution CIR for a case of 2× bandwidth enhancement. The ground-truth high-resolution CIR is shown in the same figure for comparison.

$$\begin{aligned} L(x,y) &= L_F(x,y) + L_A(x,y) \\ &= \lambda_1 \|FG_A(x) - Fy\|^2 + \lambda_2 \||G_A(x)|^2 - |y|^2\|^2 \end{aligned} \quad (7)$$

We use the mean squared error loss of ToA (eq.8) to train Regressor A, which directly estimates the coarse ToA based on the generated high-resolution CIR.

$$L(x) = \|R_A(x) - \tau_0\|^2 \quad (8)$$

After we finish the training of Regressor A, we prepare the dataset for Regressor B training. Using the ToA estimation from the Regressor A, we cycle-shift and crop each generated high resolution CIR to 1/4 length for stage B training. The task of Regressor B to minimize the difference between coarse estimation and ground truth using the training loss function (eq.9). The final ToA estimation is obtained by the summation of $[\hat{\tau_0}^{coarse}]$ and $R_B(x)$. From simulation data (Section IV), we observed Regressor B is particularly useful when we deal with high-SNR scenarios.

$$L(x) = \|R_B(x) - (\tau_0 - [\hat{\tau_0}^{coarse}])\|^2 \quad (9)$$

IV. Experiments

We train our neural networks using simulated CIRs following the channel model described in Section IV.A.1. We then conducted evaluation (without re-training) using the 802.15.4a channel model [16] and real-world measurement data. The channel model for training contains multiple parameters and we swept those parameters over a wide range to cover various channel conditions. Instead of training each neural network for a specific SNR condition, we train our neural networks only for two scenarios: high SNR ($10-30$dB) and low SNR ($0-10$dB). Proper network model selection (high vs. low SNR) is made based on SNR estimation.

We compare our proposed algorithm with MUSIC, ESPRIT and three other neural network structures; 1) a single-stage structure where only Regressor A is used without the generator and Regressor B, 2) a single-stage structure where Regressor A and B are used but the generator is replaced by interpolation, and 3) a two-stage structure where the generator and Regressor A are used and Regressor B is excluded.

A. Neural Network Training

1) Channel Model: According to (1), the channel impulse response is determined by the number of paths L, the time delay τ_l, and complex-valued attenuation factor c_l associated with l_{th} path. In our model, we assume τ_l follows a Poisson distribution with an arrival rate λ and $|c_l|$ follows a Rayleigh distribution parameterized by σ and random phase $\angle c_l$ which has a uniform distribution over $[-\pi, \pi)$. The expected power of each path is modelled as $E[|c_l|^2] = 2\sigma^2 = e^{(-\tau_l/\tau_{rms})}$ where τ_{rms} is fixed to be $75ns$. In this model, the distribution of channel is controlled by L and λ. For neural network training set generation, we randomly pick L within the range of $[3, 15]$ and λ within $[5, 50]ns$ respectively for each channel instance. Each channel is associated with a SNR which is also randomly chosen from the range of $[10, 30]dB$ for the high SNR neural network and the range of $[0, 10]dB$ for the low SNR case.

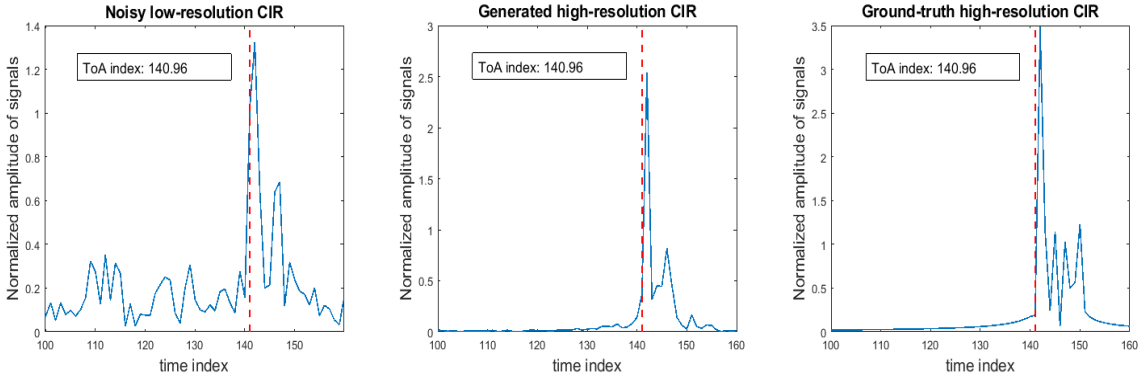

Fig. 2. Comparison of noisy low-resolution CIR (0dB), generated high-resolution CIR and ground-truth high-resolution CIR.(Red lines:the ground truth ToA.)

2) Training Details: Following the channel model, we generate a total number of $200,000$ simulated CIRs for each scenario as the training set. All the three neural networks are trained with 200 epochs and using the Adam optimizer [17] with betas = (0.9,0.999). The learning rates start with 5e-4 for generator and 1e-3 for Regressor A and B with an exponential decay factor of 0.3 for every 40 epoch.

B. Test on Simulated CIR

First, we examine the impact of L and λ to the ToA estimation problem. For each pair of L and λ, we generate 5000 testing CIRs with a fixed SNR of 20dB. The bandwidth is fixed to be 40MHz with an OFDM based CIR measurement with subcarrier spacing of 312.5kHz. Interpolation and/or the super-resolution generator increases the effective bandwidth to 80MHz. The ToA estimation is converted to the distance (d) by the relationship $d = c \times ToA$ where c is the speed of light. The distance root-mean-square error (RMSE) for ESPRIT and the proposed method are shown in Fig. 3 (darker the better) for various L and λ combinations. It is observed that the proposed method achieves consistently more reliable results regardless of L and λ while the RMSE tends to be smaller with fewer multipaths and larger multipath arrival gaps.

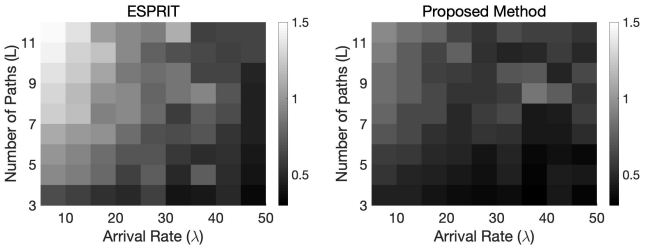

Fig. 3. Distance RMSE results for various channel parameters L and λ. (a) ESPRIT, (b) Proposed method. The darker color indicates smaller error.

Fig. 4 shows the RMSE performance for the training channel model and 802.15.4a CH1 channel, where 5000 CIRs are generated for each SNR condition. The bandwidth and subcarrier space remain the same as before. The proposed two-stage method outperforms both MUSIC and ESPRIT especially in low SNR cases where estimating ToA is more challenging. Furthermore, we evaluate the false detection rate for both channel models, where the false detection rate is defined as:

$$FD = \frac{1}{M}\sum_{i=1}^{M} I(|\hat{\tau_0}^{(i)} - \tau_0^{(i)}| > \frac{1}{2B}) \tag{10}$$

where $I(.)$ denotes the indicator function. The result in Fig. 5 shows the proposed approach exhibits lower false detection rates compared to ESPRIT and MUSIC for both channel models.

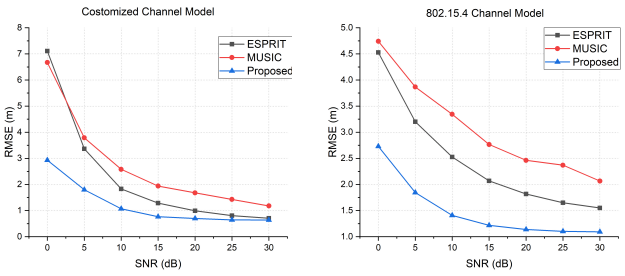

Fig. 4. RMSE performance comparison among different methods for (a) training channel model, (b) 802.15.4a CH1 channel

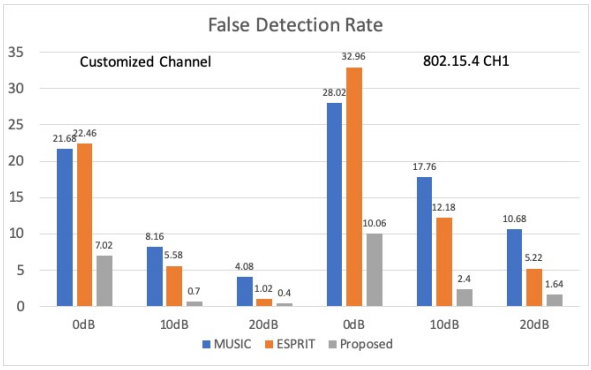

Fig. 5. False detection rate performance comparison among different methods for different channel models

Performance of the proposed neural network structure is compared with other neural network structures in table I. All

RMSE results are evaluated using the training channel model. Denoting the generator as G and the two regressors as RA and RB respectively, our proposed structure is represented by G+RA+RB and we test three other structures for comparison: RA, G+RA and RA+RB. The benefit of using the generator network is quantified by the gain from RA to G+RA and from RA+RB to G+RA+RB. Note that, for a setting without the generator G, interpolation is used and subsequent stages are retrained with interpolation results. The benefit of using a second regressor (Regressor B) is quantified by the gain from RA to RA+RB and from G+RA to G+RA+RB.

TABLE I
RMSE FOR VARIOUS NEURAL NETWORK MODELS

SNR (dB)	10	15	20	25	30
RA (Naive CNN)	1.511	1.166	1.084	1.049	1.047
G+RA (Coarse est. only)	1.229	1.004	0.932	0.903	0.901
RA+RB (Two regressors)	1.132	0.843	0.773	0.741	0.734
G+RA+RB (Proposed)	**1.071**	**0.765**	**0.698**	**0.644**	**0.638**

C. Test on Real-World Channel

Finally, we evaluate the performance of the proposed approach using the real-world measurement data. The CIR measurement is conducted using the setup in [2] to collect CIR in a building on a university campus. SNR is uncontrolled in the CIR measurement data collection and the distance for CIR measurement ranges from 20m to 40m, including non-line-of-sight multipaths. The measurement bandwidth is 80MHz and the generator increases the effective bandwidth to 160MHz. Neural networks are trained with simulated CIR data only without using the actual measured CIR. Fig. 6 compares the empirical CDF of absolute ToA error between the proposed method and ESPRIT/MUSIC, exhibiting improvement of 1.3 m in distance error at 90 percentile.

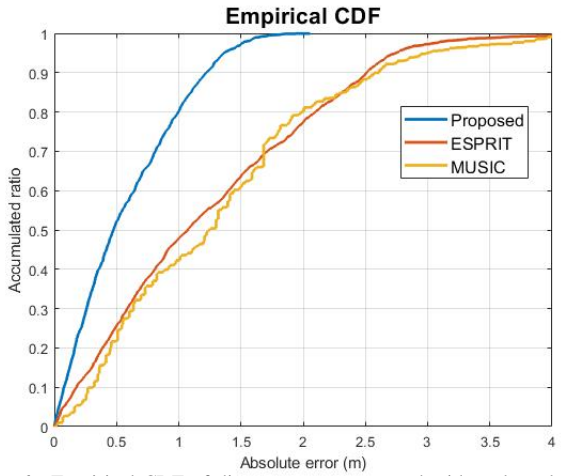

Fig. 6. Empirical CDF of distance error measured with real-world CIR.

V. CONCLUSION

In this paper we present a deep learning-based ToA estimation algorithm that utilizes a super-resolution generator to extend the effective bandwidth for ToA estimation. We propose a two-stage training and estimation process to achieve significant gains over conventional and other neural network based ToA estimation algorithms. Evaluation with real-world CIR measurements confirmed that the proposed method can reliably transfer from simulated training data to realistic channels.

REFERENCES

[1] Li-Xuan Chuo, Zhihong Luo, Dennis Sylvester, David Blaauw, and Hun-Seok Kim, "Rf-echo: A non-line-of-sight indoor localization system using a low-power active rf reflector asic tag," in *Proceedings of the 23rd Annual International Conference on Mobile Computing and Networking*, New York, NY, USA, 2017, MobiCom '17, pp. 222–234, ACM.

[2] M. Yang, L. Chuo, K. Suri, L. Liu, H. Zheng, and H. Kim, "ilps: Local positioning system with simultaneous localization and wireless communication," in *IEEE INFOCOM 2019 - IEEE Conference on Computer Communications*, April 2019, pp. 379–387.

[3] Ralph Schmidt, "Multiple emitter location and signal parameter estimation," *IEEE transactions on antennas and propagation*, vol. 34, no. 3, pp. 276–280, 1986.

[4] Richard Roy and Thomas Kailath, "Esprit-estimation of signal parameters via rotational invariance techniques," *IEEE Transactions on acoustics, speech, and signal processing*, vol. 37, no. 7, pp. 984–995, 1989.

[5] Xinrong Li and Kaveh Pahlavan, "Super-resolution toa estimation with diversity for indoor geolocation," *IEEE Transactions on Wireless Communications*, vol. 3, no. 1, pp. 224–234, 2004.

[6] Harri Saarnisaari, "Tls-esprit in a time delay estimation," in *1997 IEEE 47th Vehicular Technology Conference. Technology in Motion*. IEEE, 1997, vol. 3, pp. 1619–1623.

[7] Badri Narayan Bhaskar, Gongguo Tang, and Benjamin Recht, "Atomic norm denoising with applications to line spectral estimation," *IEEE Transactions on Signal Processing*, vol. 61, no. 23, pp. 5987–5999, 2013.

[8] Emmanuel J Candès and Carlos Fernandez-Granda, "Towards a mathematical theory of super-resolution," *Communications on pure and applied Mathematics*, vol. 67, no. 6, pp. 906–956, 2014.

[9] David Humphrey and Mark Hedley, "Super-resolution time of arrival for indoor localization," in *2008 IEEE International Conference on Communications*. IEEE, 2008, pp. 3286–3290.

[10] Gautier Izacard, Brett Bernstein, and Carlos Fernandez-Granda, "A learning-based framework for line-spectra super-resolution," in *2019 IEEE International Conference on Acoustics, Speech, and Signal Processing, ICASSP 2019 - Proceedings*, 5 2019, pp. 3632–3636.

[11] Alexander Toshev and Christian Szegedy, "Deeppose: Human pose estimation via deep neural networks," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2014, pp. 1653–1660.

[12] Morteza Mardani, Enhao Gong, Joseph Y Cheng, Shreyas S Vasanawala, Greg Zaharchuk, Lei Xing, and John M Pauly, "Deep generative adversarial neural networks for compressive sensing mri," *IEEE transactions on medical imaging*, vol. 38, no. 1, pp. 167–179, 2018.

[13] Volodymyr Kuleshov, S. Zayd Enam, and Stefano Ermon, "Audio super-resolution using neural networks," in *5th International Conference on Learning Representations, ICLR 2017, Toulon, France, April 24-26, 2017, Workshop Track Proceedings*, 2017.

[14] Alice Lucas, Michael Iliadis, Rafael Molina, and Aggelos K Katsaggelos, "Using deep neural networks for inverse problems in imaging: beyond analytical methods," *IEEE Signal Processing Magazine*, vol. 35, no. 1, pp. 20–36, 2018.

[15] Yi Sun, Xiaogang Wang, and Xiaoou Tang, "Deep convolutional network cascade for facial point detection," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2013, pp. 3476–3483.

[16] Andreas F Molisch, Kannan Balakrishnan, Chia-Chin Chong, Shahriar Emami, Andrew Fort, Johan Karedal, Juergen Kunisch, Hans Schantz, Ulrich Schuster, and Kai Siwiak, "Ieee 802.15. 4a channel model-final report," *IEEE P802*, vol. 15, no. 04, pp. 0662, 2004.

[17] Diederik P Kingma and Jimmy Ba, "Adam: A method for stochastic optimization," *arXiv preprint arXiv:1412.6980*, 2014.