Unveiling Cellular Antenna Orientations from Large Crowdsourced Datasets: A Deep Learning Approach

Lukas Eller, Philipp Svoboda, Markus Rupp
Institute of Telecommunications
TU-Wien, Vienna, Austria
{firstname}.{lastname}@tuwien.ac.at

***Abstract*—The accurate and reliable localization of transmitter locations from crowdsourced measurements has enabled the large-scale analysis of the previously hidden network layout. Recent work has shown that signal-strength measurements from drive-test campaigns also unveil the antenna orientations — providing an open-source network twin that can act as the backbone of coarse user-equipment positioning, operator benchmarking, or the generation of coverage and performance maps. In this work, we extent this drive-test based scheme to the regime of large-scale crowdsourced datasets and conduct an assessment of orientation inference on 4,950 sectors from a live LTE network. Fusing signal-strength and geometry-based features in a probabilistic Deep Learning image processing framework tackles the challenging characteristics of such noisy crowdsourced data collected in uncontrolled conditions. We further use transfer learning and weight sharing to extend our approach to allow for joint inference of sectors mounted onto the same base station. On the test dataset — representative of a complete network deployment — our selective predictors achieve median errors as low as 7.3° with 95 percentiles below 21°.**

***Index Terms*—LTE, 5G, basestation, crowdsourcing, deep learning, network topology inference, sector orientation**

I. Introduction

Since the advent of crowdsourcing, there has been increased interest in using network traces to benchmark Mobile Network Operators (MNOs) and to assess the provided quality of service from the end-user's perspective [1]. Such data have also been used to unveil the typically hidden network infrastructure by constructing public databases of transmitter locations [2], [3] — be it as a backbone for user equipment (UE) positioning [4] or to identify un- or under-served regions [5], [6]. Further, such public network twins can power system level simulations [7], or be used to construct performance and throughput maps by combining rate prediction models [8] with spatial regression scheme, requiring knowledge of the underlying infrastructure [9]. While base station (BS) localization has been actively studied, inference of the sectorization has so far only been conducted on clean drive-test measurements [10] — which we find to be not directly comparable to inherently noisy crowdsourced datasets [1], [4]. In fact, due to the challenging data characteristics, BS localization has also long relied on rather simple heuristics [2]. The 4&5G Timing Advance (TA) parameter does, however, offer a suitable indicator for the distance between UE and transmitter and allows for reliable BS localization [3], [11]. In this work, we build on the first orientation inference scheme assessed in [10] and adapt it for a large scale application on noisy crowdsourced data. In particular, we:

i) Treat the inference problem in a deep learning (DL) image processing framework which allows for the effective fusion of geometric and signal-strength approaches — tackling the uncontrolled measurement conditions of crowdsourcing. A probabilistic formulation through a Von-Mises Biternion network ensures that unreliable data instances are identified [12]. *ii*) Conduct the first large-scale analysis of sector orientation inference using a crowdsourced dataset provided by an Austrian MNO consisting of Reference Signal Received Power (RSRP) measurements and their relative positions to the BS. With ground-truth orientations for 4,950 LTE sectors spread all over Austria, the dataset accounts for around 40% of the operator's cells in the considered band. *iii*) Utilize transfer learning and weight sharing to exploit the dependencies of sectors mounted on the same BS such that no complete retraining is required.

II. Related Work

While sectorization inference has so far only been applied to a controlled drive-test setup [10], there is a long history of localizing BS through crowdsourced measurements [2]. Because of the often noisy data characteristics and limited availability of parameters [1], these estimates have typically relied on rather simple heuristics — such as the centroid of all measurement locations weighted by the received signal-strength [2]. It is still possible to obtain reliable locations through the 4&5G TA parameter or by accounting for unlabeled cell-tower locations, which both have no equivalent for the orientation inference use case considered in this work [3], [11]. Here, the authors' previously proposed Bayesian scheme from [10] accounting for shadow fading statistics showed promising results – with median errors of 18° for the single-sector and 11° for the joint sector approach. Unfortunately, we were not able to directly translate these results to the crowdsourcing context in our experiments. In fact, applying the model based approach from [10] to our noisy dataset resulted in frequent failures of the Markov Chain Monte Carlo (MCMC) sampler — also due to the challenge of specifying the pathloss model for a noisy country-wide dataset and a heterogeneous set of devices. This motivated

978-1-6654-6975-3/22/$31.00 ©2022 IEEE

an extension of the original approach to the big-data regime, such that it can, in addition to the RSRP, also take advantage of the geometry induced by the cell assignment, which is often exploited for BS localization [2], [13]. Further, we abstain from explicit pathloss modelling but replace it with a data-driven DL approach [14], which can better adapt to the noisy characteristics of the crowdsourcing setting. Still, we keep major aspects of the original approach from [10] — in particular, representing the angular output together with an associated confidence through a Von-Mises Biternion network [12]. This way, we can assess the performance in a selective prediction framework and identify unreliable data instances [3], [15]. Overall, we follow a similar approach as the localization scheme in [11], which encodes geometric, TA and signal-strength features through an image processing approach. A major advantage of this image formulation is, that the inputs have equal sizes regardless the number of measurements per sector, which makes them suitable for convolutional neural networks (CNNs) [14]. This setup also enables extensive data-augmentation, which combats overfitting and boosts generalization [16]. We further translate the non-overlapping sector assumption from [10] to the DL framework. In particular, we exploit these inter-sector dependencies by jointly processing three-sector BSs through a dedicated composite model inspired by Siamese networks [17]. Using pretrained single-sector processing networks in a transfer learning manner further ensures that no complete retraining is required [18].

In the following we will first introduce the image generation process, before we detail the loss function and network architecture of our model.

III. Image Generation & Data Augmentation

For a single sector with ground-truth orientation $\phi_{\text{sec}} \in [0, 2\pi)$ we retrieve the raw measurement set

$$\mathcal{M} = \{\mathbf{X}_{\text{pos}}, \mathbf{r}_{\text{meas}},\ \phi_{\text{sec}}\}, \tag{1}$$

with the recorded RSRP values collected into the vector $\mathbf{r}_{\text{meas}}$ of length M and the corresponding positions relative to the serving BS stacked in the matrix $\mathbf{X}_{\text{pos}} \in \mathbb{R}^{M\times 2}$. During processing we *i*) generate fixed sized images out of the inputs through spatial binning and *ii*) conduct data augmentation to increase the generalization capabilities of our models [16].

A. Image Generation

We generate a fixed sized two channel image tensor $\mathbf{X}_{\text{img}}^{(i)}$, consisting of the sample mean and standard deviation over binned subsets of the RSRP measurements for the given sector:

$$f_{\text{img}}(\mathbf{r}_{\text{meas}}^{(i)},\ \mathbf{X}_{\text{pos}}^{(i)}) \ \rightarrow\ \mathbf{X}_{\text{img}}^{(i)} \in \mathbb{R}^{K\times K\times 2}. \tag{2}$$

The $K \times K$ uniformly spaced bins are thereby constructed through a spatial rectangular grid with boundary distance d_{lim}, so that the bin spacing in both directions is given by $d_{\text{bin}} = \frac{2d_{\text{lim}}}{K}$[1]. Normalizing the two channels to the interval $[0, 1]$ allows to represent empty bins by setting the corresponding entries of $\mathbf{X}_{\text{img}}$ to -1. Overall, this results in an input representation with a fixed dimension and enables pattern recognition using CNNs.

[1]We use $K = 130$ and $d_{\text{lim}} = 2000\ m$, resulting in $d_{\text{bin}} \approx 30\ m$.

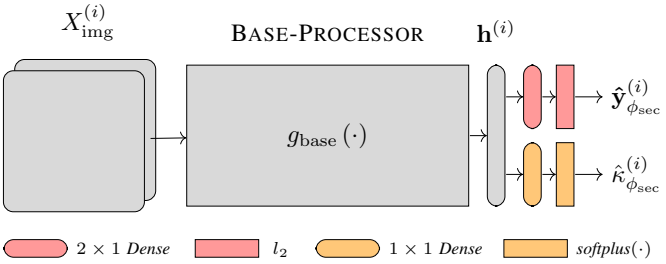


Fig. 1: Architecture of the used Von-Mises Biternion network.

B. Data Augmentation

During training, we augment $\mathcal{M}$ and construct artificial training samples $\mathcal{M}_{\text{rot}}^{'} = \{\mathbf{X}_{\text{pos}}^{'}, \mathbf{r}_{\text{meas}}, \phi_{\text{sec}}^{'}\}$ by rotating the measurement with respect to the BS location. Hence, we apply a suitable rotation matrix $\mathbf{R}(\alpha)$ to the measurement positions $\mathbf{X}_{\text{pos}}$ and adapt the target values ϕ_{sec} accordingly. In a second step, we generate augmented input samples $\mathcal{M}_{\text{sub}}^{'} = \{\mathbf{X}_{\text{pos}}^{'}, \mathbf{r}_{\text{meas}}^{'}, \phi_{\text{sec}}\}$ through subsampling to better account for possibly sparse inputs. As such, we uniformly at random select an integer number $M^{'} \sim \mathcal{U}(1, M)$ of measurements from $\mathcal{M}$ without replacement. Overall, these augmentation steps significantly reduce overfitting and help to improve generalization.

IV. Problem Statement & Loss Formulation

Given an input sample $\mathbf{X}_{\text{img}}$, we intend to predict the sector orientation $\hat{\phi}_{\text{sec}}$ together with a measure of confidence $\hat{\kappa}_{\phi_{\text{sec}}}$ through a neural network (NN) with weight vector $\boldsymbol{\theta}$.

We thus treat our network outputs $\hat{\phi}_{\text{sec}}$ and $\hat{\kappa}_{\phi_{\text{sec}}}$ as a parameterization of the Von-Mises distribution:

$$\mathcal{VM}(\phi_{\text{sec}}; \hat{\phi}_{\text{sec}}, \hat{\kappa}_{\phi_{\text{sec}}}) = \frac{e^{\hat{\kappa}_{\phi_{\text{sec}}} \cos(\phi_{\text{sec}} - \hat{\phi}_{\text{sec}})}}{2\pi I_0(\hat{\kappa}_{\phi_{\text{sec}}})}, \tag{3}$$

which can be interpreted as the circular equivalent to a Gaussian distribution, with $I_0(\cdot)$ denoting the modified Bessel function of the first kind and order zero [12].

Angular regression tasks are circular by nature [19]. Hence, we first translate the targets ϕ_{sec} into their respective Biternion representations $\mathbf{y}_{\phi_{\text{sec}}} = \begin{bmatrix} \cos(\phi_{\text{sec}}) & \sin(\phi_{\text{sec}}) \end{bmatrix}$ [12], before minimizing the negative log-likelihood of the Von-Mises distribution over the N_{train} training samples:

$$\begin{aligned} \mathcal{J}(\boldsymbol{\theta}) &= -\sum_{i=1}^{N_{\text{train}}} \log \mathcal{L}_{\text{VM}}\left(\phi_{\text{sec}}^{(i)};\ \mathbf{X}_{\text{img}}^{(i)},\ \boldsymbol{\theta}\right) \\ &= -\sum_{i=1}^{N_{\text{train}}} \hat{\kappa}_{\phi_{\text{sec}}}^{(i)} \cdot \left\langle \mathbf{y}_{\phi_{\text{sec}}}^{(i)}, \hat{\mathbf{y}}_{\phi_{\text{sec}}}^{(i)} \right\rangle \\ &\quad + \sum_{i=1}^{N_{\text{train}}} \log\left(2\pi I_0\left(\hat{\kappa}_{\phi_{\text{sec}}}^{(i)}\right)\right). \end{aligned} \tag{4}$$

Note, that in (4) the $\cos(\cdot)$ from (3) was replaced by the inner product of $\mathbf{y}_{\phi_{\text{sec}}}$ and $\hat{\mathbf{y}}_{\phi_{\text{sec}}}$, denoting the unit-norm Biternion representations of ϕ_{sec} and $\hat{\phi}_{\text{sec}}$ [12].

The overall model architecture is sketched in Fig. 1. We first pass the inputs $\mathbf{X}_{\text{img}}$ through a generic BASE-PROCESSOR network, suitable for image-processing tasks, to generate the intermediate representation $\mathbf{h} = g_{\text{base}}(\mathbf{X}_{\text{img}})$ with $\mathbf{h} \in \mathbb{R}^{N_h}$. A set of subsequent layers then outputs the final vector estimate $\hat{\mathbf{y}}_{\phi_{\text{sec}}}$ together with a measure of confidence $\hat{\kappa}_{\phi_{\text{sec}}}$. In Fig. 1, the application of the differentiable l_2 normalization layer to a two neuron output enforces the unit-norm constraint for $\hat{\mathbf{y}}_{\phi_{\text{sec}}}$, resulting in the expression

$$\hat{\mathbf{y}}_{\phi_{\text{sec}}} = \frac{\mathbf{W}\mathbf{h} + \mathbf{b}}{\|\mathbf{W}\mathbf{h} + \mathbf{b}\|_2}, \tag{5}$$

with $\mathbf{W} \in \mathbb{R}^{2 \times N_h}$ and $\mathbf{b} \in \mathbb{R}^{2 \times 1}$ [12]. At the same time, we generate the concentration estimate $\hat{\kappa}_{\phi_{\text{sec}}}$ through a softplus activation function to map the output of a single neuron to $\mathbb{R}^+$. In particular,

$$\hat{\kappa}_{\phi_{\text{sec}}} = \log\left(1 + e^{\left(\mathbf{w}^T \mathbf{h} + b\right)}\right), \tag{6}$$

with $\mathbf{w} \in \mathbb{R}^{N_h}$ and $b \in \mathbb{R}$. We refer the interested reader to Appendix A for a detailed description of the selected BASE-PROCESSOR network configuration used to generate the intermediate representation $\mathbf{h}$.

With the model architecture specified, we now extend the framework from Fig. 1 to account for the dependencies between orientations of sectors mounted onto the same BS.

V. JOINT SECTOR EXTENSION & WEIGHT SHARING

In [10], we achieved a significant error reduction by accounting for typical eNodeB (eNB) configurations — in the simplest case the sectors are assumed to be non-overlapping. By jointly processing samples from the same eNB we can exploit such dependencies and also account for patterns in the assignment of UEs among the individual sectors.

The proposed composite model — shown for the three-sector case in Fig. 2 — consists of stacked BASE-PROCESSOR networks from the SINGLE-SECTOR use case, with the concatenated representations $\mathbf{h}^{(i)}$ from Fig. 1 acting as the input to a second COMBINER network. This COMBINER builds a joint representation

$$\mathbf{h}^{(i)}_{\text{joint}} = g_{\text{comb}}\left(\mathbf{h}^{(i)}_1, \mathbf{h}^{(i)}_2, \mathbf{h}^{(i)}_3\right), \tag{7}$$

from which the individual per-sector $\hat{\mathbf{y}}_{\phi_{\text{sec}}}$ and $\hat{\kappa}_{\phi_{\text{sec}}}$ estimates are constructed with dedicated l_2 and softplus layers. Inspired by Siamese networks [17], we directly account for the symmetry of the stacked composite model, by linking the parameters of the individual processing networks through weight sharing, so that they are optimized jointly. This reduces the total number of trainable parameters by a factor of three. Clearly, this composite model can only be trained

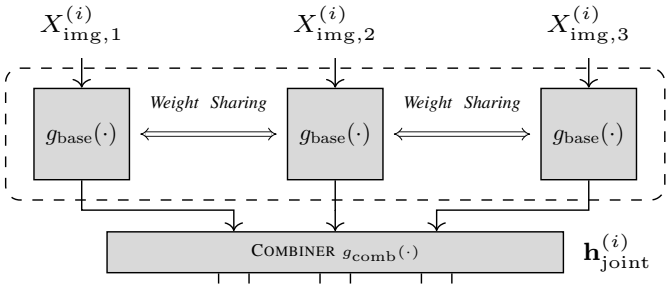


Fig. 2: Composite model generating joint representation $\mathbf{h}_{\text{joint}}$.

on the three-sector subset of the training data. Indirectly, we can still utilize all available data by reusing the pretrained BASE-PROCESSOR networks from the SINGLE-SECTOR case and dedicating the THREE-SECTOR training phase to the optimization of the newly added COMBINER. As common in transfer learning, we still allow for some fine-tuning of the processing networks, but with a significantly smaller learning rate than for $\mathbf{g}_{\text{comb}}(\cdot)$ [18]. The detailed model configuration is again provided in Appendix A.

Overall, the combination of weight sharing and transfer learning ensures that the composite model can extract common BS configurations and account for the cell assignment among sectors, without requiring complete retraining of the BASE-PROCESSOR networks. Indirectly, it can thus make use of the complete training dataset.

VI. SELECTIVE PREDICTION

When assessing the performance of our model, we specifically address the provided measure of uncertainty which allows us to conduct selective prediction [15]. Hence, we abstain from prediction whenever the provided confidence is not sufficient. Based on that, we first consider the selection function g $(\cdot)$, which accepts estimates whenever the concentration threshold t_κ is fulfilled:

$$g\left(\hat{\kappa}_{\phi_{\text{sec}}}, t_\kappa\right) = \begin{cases} 1, & \hat{\kappa}_{\phi_{\text{sec}}} \geq t_\kappa, \\ 0, & \text{else} \end{cases}. \tag{8}$$

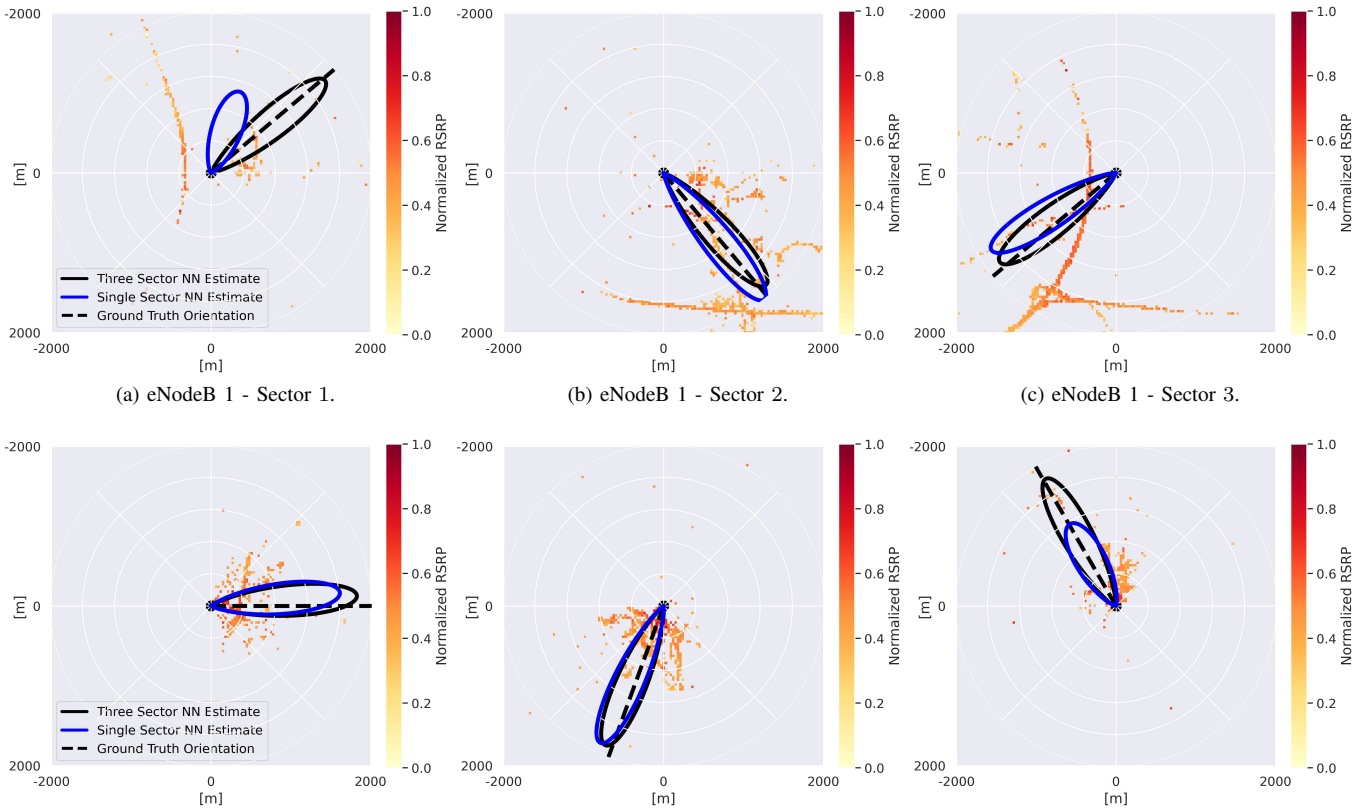
The notion of coverage $c(\cdot)$ is then defined as the ratio of predictions from the test dataset which are accepted by $g\,(\cdot)$:

$$c(t_\kappa) = \frac{1}{N_{\text{test}}} \sum_{i=1}^{N_{\text{test}}} g\left(\hat{\kappa}^{(i)}_{\phi_{\text{sec}}}, t_\kappa\right). \tag{9}$$

For a given threshold t_κ, we can then report the mean error for all accepted estimates together with the resulting value of $c(\cdot)$ [15]. In the following evaluation, we will use this terminology to relate the prediction error to the provided confidence in a unified framework.

VII. REAL WORLD PERFORMANCE EVALUATION

Overall, the crowdsourced dataset provided by an Austrian MNO consists of measurements from 4,950 LTE sectors transmitting at 800 Mhz with locations distributed all around Austria. For each sector, the MNO validated the ground-truth



(a) eNodeB 1 - Sector 1. (b) eNodeB 1 - Sector 2. (c) eNodeB 1 - Sector 3.

(d) eNodeB 2 - Sector 1. (e) eNodeB 2 - Sector 2. (f) eNodeB 2 - Sector 3.

Fig. 3: Exemplary predictions for two Three-Sector eNBs, with the subplots showing the measurements for each sector.

orientations ϕ_{sec} and the corresponding eNB location. We retrieve up to 1,000 randomly selected measurements for each sector — resulting in a total of $\approx 2,600,000$. Out of the 4,950 sectors, 2,658 sectors belong to fully represented three-sector eNBs.

To allow for comparison of the SINGLE- and THREE-SECTOR use cases, we construct the test dataset solely of randomly selected three-sector eNBs, which we treat as individual sectors for the SINGLE-SECTOR approach. With an overall test ratio of around 15%, this results in a test dataset consisting of $N_{\text{test}} = 750$ sectors and a training dataset consisting of $N_{\text{train}} = 4,200$ sectors for the SINGLE-SECTOR and $N_{\text{train}} = 1,908$ for the subsequent fine-tuning of the THREE-SECTOR setup.

A. Performance for exemplary Three-Sector eNodeBs

Fig. 3 shows exemplary predictions of the trained model for two different three-sector eNB from our test dataset. Here, each subplot shows the mean RSRP channel of $\mathbf{X}_{\text{img}}$ for the given sector together with a polar representation of the Von-Mises output parameterized on $\hat{\phi}_{\text{sec}}$ and $\hat{\kappa}_{\phi_{\text{sec}}}$. For all predictions, we observe clear patterns in the measurement distribution; in particular outlines of roads passing through Figs. 3a-3c. When comparing the SINGLE- and THREE-SECTOR approach, we find that the latter is generally more confident in it's predictions and achieves a lower error overall. For both eNBs, the THREE-SECTOR approach reliably fuses the BASE-PROCESSOR representations and accounts for frequent BS configurations as well as the UE assignment among sectors. This is apparent in Fig. 3a, where it correctly associates the measurements with a 90° offset configuration commonly found in our dataset. Meanwhile, the SINGLE-SECTOR NN is aware of the uncertainty and reports a significantly lower concentration parameter $\hat{\kappa}_{\phi_{\text{sec}}}$ for the given estimate.

B. Performance for complete Test Dataset

For the overall evaluation we consider the baselines from [10], which are angular estimators derived from widely used heuristics for BS localization [2]. In particular, we include the ANGULAR MEAN estimator which exploits the cell assignment and reports the angular mean of all measurements for a given sector [10]. We further assess the MAX RSRP estimator, reporting the relative angle of the measurement with the highest signal-strength. To compare the performance in the selective prediction framework we interpret M, the number of mesurements per sector, as a measure of confidence — together with an associated threshold t_M.

Fig. 4 depicts the prediction errors for the proposed estimators and the considered baselines. In accordance with VI, we report the errors over the complete test dataset —

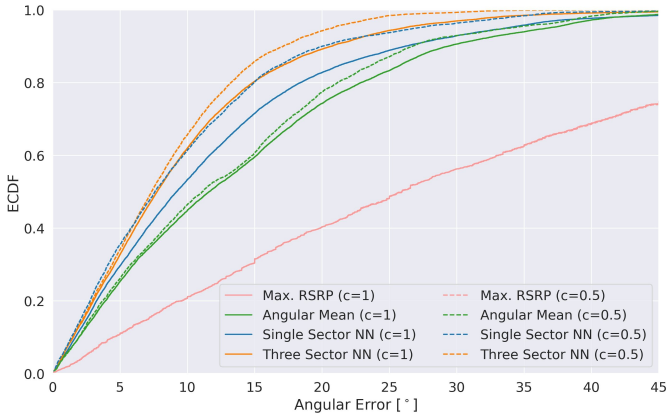


Fig. 4: Prediction errors for all sectors in test dataset $c(\cdot) = 1$ vs. results for more confident half of predictions $c(\cdot) = 0.5$.

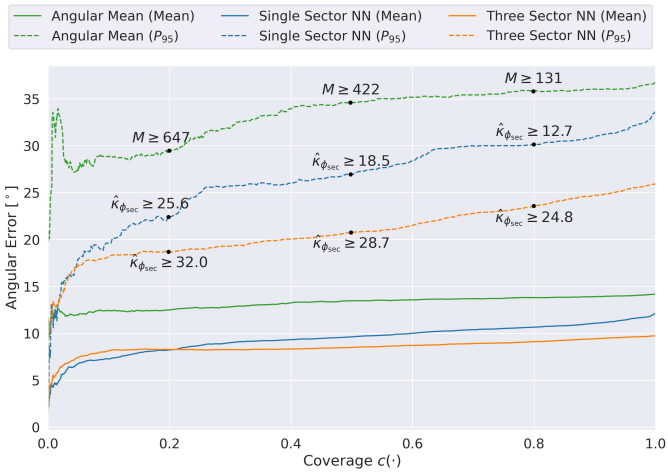


Fig. 5: Mean and 95 percentile of the error plotted over ratio of remaining samples $c(\cdot)$ — annotated with selected thresholds.

$c(\cdot) = 1$ — and additionally show the performance for the confident half of the predictions[2] — $c(\cdot) = 0.5$. For $c(\cdot) = 1$ we note, that the THREE-SECTOR and SINGLE-SECTOR NNs perform best with median errors of $7.75°$ and $9.19°$, followed by the ANGULAR MEAN with $11.62°$ and the MAX RSRP baseline with $25.91°$. Clearly, the proposed estimators outperform the $11°$ and $18°$ of the purely signal-strength based approach from [10], although these were obtained on a clean drive-test dataset. The apparent difference between the two baselines in Fig. 4 already highlights the importance of geometric features for the crowdsourced dataset — while the MAX RSRP falls short, the ANGULAR MEAN estimator utilizes the geometric information from the cell-assignment, that was not available for the scanner measurements from [10]. The errors for the confident half of predictions with $c(\cdot) = 0.5$ show, that both the SINGLE- and THREE-SECTOR NN provide a reliable measure of confidence — reducing the median error to $7.58°$ and $7.34°$ respectively. This is only partially the case for the ANGULAR MEAN estimator with $11.11°$, while the MAX RSRP does not seem to benefit from removing sparse measurements at all.

C. Assessment of Estimator Confidence

To assess the measures of confidence in more detail we further compute the angular error together with the associated coverage $c(\cdot)$ for all possible threshold values. The resulting Fig. 5, termed a risk-coverage curve in the literature [15], shows both the mean and 95 percentile of the error. We find, that the errors are decreasing for higher values of the thresholds — demonstrating that we can reliable trade off prediction error vs. coverage $c(\cdot)$. While the THREE-SECTOR network performs well for $c(\cdot) = 1$, the SINGLE-SECTOR NN significantly reduces the number of outliers as well as the mean error. The annotated thresholds in Fig. 5 also indicate, that the THREE-SECTOR approach does generally exhibit a higher confidence than the SINGLE-SECTOR counterpart, which seems justified given the lower overall error.

[2]Thresholds for $c(\cdot) = 0.5$ are given by the median of the confidence.

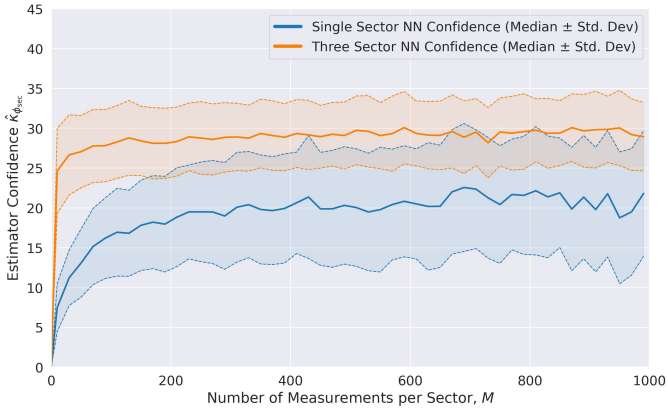


Fig. 6: The NN confidence is learned from the training data and depends only partially on the measurement count M.

Fig. 6, showing SINGLE- and THREE-SECTOR estimator confidence $\hat{\kappa}_{\phi_{\text{sec}}}$ over the measurement count M, further displays this relation. We observe, that the confidence measures provided by the NNs only partially depend on M, indicating that they indeed account for RSRP and user distribution when computing the confidence estimate from the input $\mathbf{X}_{\text{img}}$. For the SINGLE-SECTOR estimator, the confidence increases with M until it reaches a plateau starting from $M > 200$. Meanwhile, the THREE-SECTOR estimator is rather independent on the number of samples in a given sector. This is not surprising, as in most cases it can compensate for sparse measurements by relying on the data from the two remaining sectors, which already allow for a reasonable estimate for the complete eNB.

VIII. CONCLUSION

In this work we show, that the orientation of LTE sector antennas can be estimated reasonably well from large, noisy crowdsourced datasets. In our experiments, we found model-based approaches to be inappropriate for this challenging regime. We thus reformulated the angular regression problem into a DL based image processing task with a dedicated Biternion output and the Von-Mises log-likelihood as a loss

function — effectively fusing signal-strength and geometry-based features. The evaluation on real world crowdsourced measurements from a live LTE network shows, that we can reliably trade off risk for coverage, with a median error of 7.6° and a 95 percentile of 27°. The simplistic ANGULAR MEAN estimator too performs well over a large portion of the dataset — showing that the geometry induced by the cell assignment is a critical feature. By deploying a transfer-learning-based composite model, trained jointly on triplets of sectors from the same eNB, the performance of the DL models can be further improved to a median error of 7.3° and a 95 percentile of 21°. Overall, we find the purely data-driven scheme with uncertainty estimation well suited for such noisy crowdsourced datasets, given that it significantly outperforms the 11° and 18° of the model-based scheme evaluated on a dedicated drive-test campaign.

REFERENCES

[1] V. Raida, P. Svoboda, M. Lerch, and M. Rupp, "Crowdsensed performance benchmarking of mobile networks," *IEEE Access*, vol. 7, pp. 154899–154911, 2019.

[2] E. Neidhardt, A. Uzun, U. Bareth, and A. Küpper, "Estimating locations and coverage areas of mobile network cells based on crowdsourced data," in *Proc. 6th Joint IFIP Wirel. Mob. Netw. Conf.*, pp. 1–8, IEEE, 2013.

[3] L. Eller, V. Raida, P. Svoboda, and M. Rupp, "Localizing basestations from end-user timing advance measurements," *IEEE Access*, vol. 10, pp. 5533–5544, 2022.

[4] M. Ulm, P. Widhalm, and N. Brändle, "Characterization of mobile phone localization errors with opencellid data," in *2015 4th International Conference on Advanced Logistics and Transport (ICALT)*, pp. 100–104, IEEE, 2015.

[5] M.-R. Fida and M. K. Marina, "Uncovering mobile infrastructure in developing countries with crowdsourced measurements," in *Proceedings of the Tenth International Conference on Information and Communication Technologies and Development*, pp. 1–11, 2019.

[6] F. Malandrino, C.-F. Chiasserini, and S. Kirkpatrick, "Cellular network traces towards 5G: Usage, analysis and generation," *IEEE Transactions on Mobile Computing*, vol. 17, no. 3, pp. 529–542, 2018.

[7] M. K. Müller, F. Ademaj, T. Dittrich, A. Fastenbauer, B. R. Elbal, A. Nabavi, L. Nagel, S. Schwarz, and M. Rupp, "Flexible multi-node simulation of cellular mobile communications: the Vienna 5G system level simulator," *EURASIP J. Wirel. Commun. Netw.*, vol. 2018, no. 1, pp. 1–17, 2018.

[8] D. Minovski, N. Ogren, C. Ahlund, and K. Mitra, "Throughput prediction using machine learning in LTE and 5G networks," *IEEE Transactions on Mobile Computing*, pp. 1–1, 2021.

[9] J. Thrane, B. Sliwa, C. Wietfeld, and H. L. Christiansen, "Deep learning-based signal strength prediction using geographical images and expert knowledge," in *GLOBECOM 2020-2020 IEEE Global Communications Conference*, pp. 1–6, IEEE, 2020.

[10] L. Eller, P. Svoboda, and M. Rupp, "Bayesian inference of sector orientation in LTE networks based on end-user measurements," in *2021 IEEE 94th Vehicular Technology Conference (VTC2021-Fall)*, pp. 1–6, IEEE, 2021.

[11] A. Ghasemi and J. Parekh, "Deep learning based localization of LTE enodebs from large crowdsourced smartphone datasets," in *2021 IEEE 93rd Vehicular Technology Conference (VTC2021-Spring)*, pp. 1–7, IEEE, 2021.

[12] S. Prokudin, P. Gehler, and S. Nowozin, "Deep directional statistics: Pose estimation with uncertainty quantification," in *Proceedings of the European conference on computer vision (ECCV)*, pp. 534–551, 2018.

[13] S. Sundberg and J. Garcia, "Locating enodebs through sectorization inference—sector fitting evaluated on a railway use case," *Computer Networks*, vol. 190, p. 107945, 2021.

[14] I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning*. MIT Press, 2016. http://www.deeplearningbook.org.

[15] Y. Geifman and R. El-Yaniv, "Selective classification for deep neural networks," in *Proceedings of the 31st International Conference on Neural Information Processing Systems*, NIPS'17, (Red Hook, NY, USA), p. 4885–4894, Curran Associates Inc., 2017.

[16] C. Shorten and T. M. Khoshgoftaar, "A survey on image data augmentation for deep learning," *Journal of Big Data*, vol. 6, no. 1, pp. 1–48, 2019.

[17] G. Koch, R. Zemel, R. Salakhutdinov, *et al.*, "Siamese neural networks for one-shot image recognition," in *ICML deep learning workshop*, vol. 2, Lille, 2015.

[18] H.-W. Ng, V. D. Nguyen, V. Vonikakis, and S. Winkler, "Deep learning for emotion recognition on small datasets using transfer learning," in *Proceedings of the 2015 ACM on international conference on multimodal interaction*, pp. 443–449, 2015.

[19] L. Beyer, A. Hermans, and B. Leibe, "Biternion nets: Continuous head pose regression from discrete training labels," in *German Conference on Pattern Recognition*, pp. 157–168, Springer, 2015.

[20] "Tensorflow." https://www.tensorflow.org/. Accessed: 2022-01-26.

APPENDIX

We use *Tensorflow* [20] and train our networks on a *NVIDIA GeForce RTX 3060 Ti* with the *Adam* optimizer. The model configurations were obtained through manual experimentation and hyperparameter optimization with *Keras Tuner* to reach a balance between performance and complexity.

Layer / Parameter	Configuration
Conv2D Kernel Size (BASE-PROCESSOR)	$\{(5,5),(3,3),(3,3)\}$
Conv2D Filter Size (BASE-PROCESSOR)	$\{64,64,64\}$
Conv2D Activation (BASE-PROCESSOR)	{Relu, Relu, Relu}
MaxPooling2D (BASE-PROCESSOR)	$\{(2,2),(2,2),(2,2)\}$
Dense Layer Size (BASE-PROCESSOR)	$\{128,64,10\}$
Dense Layer Activation (BASE-PROCESSOR)	{Relu, Relu, linear}
Dropout Layer $p(\cdot)$ (BASE-PROCESSOR)	$\{0.2,0.2,0\}$
Learning Rate	$[1\times10^{-3}, 1\times10^{-6}]$
ReduceLROnPlateau	Factor: 0.1, Patience: 25
Batch Size	32
Epochs	150

TABLE I: SINGLE-SECTOR network configuration.

Tab. I shows the model configuration and training procedure for the SINGLE-SECTOR use case. Here, we utilize three consecutive *Conv2D* layers with *Relu* activations followed by *BatchNormalization* and *MaxPooling2D* operations [14]. The flattened output is then processed by a *Dense* network to form $\mathbf{h}^{(i)}$ with $N_h = 10$. During training, we control the learning rate through a *ReduceLROnPlateau* scheduler [20].

Layer / Parameter	Configuration
Dense Layer Size (COMBINER)	$\{64,30\}$
Dropout Layer $p(\cdot)$ (COMBINER)	$\{0.2,0\}$
Learning Rate (BASE-PROCESSOR)	1×10^{-5}
Learning Rate (COMBINER)	$[1\times10^{-3}, 1\times10^{-6}]$
ReduceLROnPlateau (COMBINER)	Factor: 0.1, Patience: 25
Batch Size	32
Epochs	50

TABLE II: THREE-SECTOR network configuration.

Tab. II shows the configuration for the THREE-SECTOR fine-tuning, where we use the pretrained BASE-PROCESSORS from Tab. I with weight sharing. The COMBINER processes the concatenated representations through a *Dense* layer using *BatchNormalization* and *Dropout*. It then outputs $\mathbf{h}_{\text{joint}}$ with $N_{\text{h,joint}} = 30$. For the COMBINER we again use a *ReduceLROnPlateau* scheduler, but set a fixed small learning rate for the fine-tuning of the BASE-PROCESSOR networks.