

Peer Review

Review of: "Deep Learning-Based CKM Construction with Image Super-Resolution"

Evgenii Arsentev¹

1. Independent researcher

1. What the work claims

This letter treats the construction of a channel knowledge map (CKM) from sparse measurements as an image super-resolution (SR) problem. Sparse measurements taken on a uniform grid play the role of the low-resolution image, and the full map is the high-resolution image (Introduction, Fig. 1; Section II, Eq. (1)). Unlike the CKM methods the authors cite (RadioUNet, cGAN-based, I2I inpainting), the approach uses no building map or other side information, only the sparse measured values (Abstract; Introduction). The model is an SRResNet with 1,549,462 parameters and five residual blocks. It is trained with an MSE loss because the goal is accurate channel values rather than perceptual realism (Section III.A, III.B). The SR factors are $k = 2, 4, 8$, and 16, which correspond to measuring $1/4$, $1/16$, $1/64$, and $1/256$ of the locations (Section III.B).

For data, the authors use two simulated datasets: RadioMapSeer (path loss) and CKMImageNet (path loss and angle-of-arrival, AoA, maps, generated with ray tracing) (Section IV.A.1). The baselines are nearest neighbor interpolation, bicubic interpolation, and SRGAN (Section IV.B.1). The headline claims are that SRResNet beats all three baselines on every metric and that with $1/16$ of the locations measured, the path loss RMSE is 1.1 dB on RadioMapSeer (Abstract; Section IV.B.1, Table I). The authors also state that the same network can build channel angle maps (CAMs) (Section IV.B.3, Fig. 7). I will comment on how well the evidence supports these claims in the sections below.

2. Whether the conclusion holds on the data

The paper is clear and easy to follow, and the idea is a useful one. The method needs only the sparse values and no building map, and the super-resolution framing makes a large computer-vision toolbox

available to CKM work (Abstract; Introduction). The numbers support SRResNet over the other methods. In Table I (RadioMapSeer, $4\times$ SR), the RMSE is 6.283 dB for nearest, 5.182 dB for bicubic, 1.409 dB for SRGAN, and 1.063 dB for SRResNet, and the PSNR is 41.53 for SRResNet against 39.82 for SRGAN. In Table II (CKMImageNet, $4\times$ SR), the RMSE is 24.97, 20.89, 13.12, and 7.684 dB in the same order, with PSNR 30.87 for SRResNet against 25.42 for SRGAN. The ranking is the same on both datasets.

On the test sets, I think 1,000 test maps per dataset is a reasonable size for this comparison (Section IV.A.1). RadioMapSeer has 20,000 train and 1,000 test maps, CKMImageNet path loss has 11,064 train and 1,000 test, and AoA has 11,866 train and 1,000 test. For RadioMapSeer, the authors ensure that the test base-station locations are not in the training set. For CKMImageNet, the split is described only as "randomly select," with no statement on separating base stations or environments. With a random split, a reader cannot tell whether maps from the same environment sit in both the training and test sets. A single sentence on this would settle it.

On baselines and statistics, the baselines are two interpolators and SRGAN. The interpolators are a low bar, so the comparison with SRGAN is the informative one. It would help to also run the learned CKM methods the paper itself cites, RadioUNet [9], Kriging [6], and I2I inpainting [11], even in a sparse-input-only setting, because a learned CKM baseline would make that comparison stronger (Introduction; Section IV.B). Every number in Tables I and II is a single value, with no standard deviation, no confidence interval, and no repeated training runs. I would add the mean and spread over several runs, or over the test maps, to both tables.

A few claims go a little beyond the evidence. The abstract headline of 1.1 dB comes from RadioMapSeer, while on CKMImageNet at the same $4\times$ factor, SRResNet has 7.684 dB (Table II). The abstract does not mention this second number, and I would add it there. AoA maps are evaluated only visually (Fig. 7), and there is no numeric metric for AoA in any table, so the claim that "SR techniques are also effective for CAM construction" (Section IV.B.3) rests on images alone. A numeric AoA metric would give it firmer support. Finally, the subsection is titled "Performance and Stability" (Section IV.B), and Section IV.B.2 says SRResNet "exhibits more stable performance compared to SRGAN". No spread across runs or maps is reported, so this statement is not yet backed by data. Reporting the spread would address it, or the wording could be softened.

3. Reproducibility

Both datasets are public, with links in footnotes 1 and 2 (radiomapseer.github.io and the CKMImageNet GitHub repository), and I appreciate that the split sizes are given in Section IV.A.1. The training details that are reported are also helpful: model size of 1,549,462 parameters (Section III.A), and batch size 32, 100,000 iterations, learning rate 0.001, with image cropping removed (Section III.B, Algorithm 1). The pixel-to-value mapping is stated clearly in Section IV.A.2: CKMImageNet path loss from -250 dB to -50 dB onto grayscale 0-255, AoA from -180 to 180 degrees onto 0-255 with buildings set to -200 degrees, and RadioMapSeer path loss from -147 dB to -47 dB. The CKMImageNet image size of 128×128 is given in Section IV.A.1.

Some details I could not find in the text. I would add a link to the code or the trained weights, the optimizer, the random seeds, and the number of training runs, and the hardware and training time. It would also help to state the training settings used for the SRGAN baseline, and the image size of the RadioMapSeer maps. The exact formula that turns pixel MSE into RMSE in dB is not given either (Section II; Tables I and II), for example, whether building pixels are excluded.

One point needs a clearer description: how the low-resolution input is made. Eq. (1) defines it as pure sampling by a 0/1 matrix H , while Section III.B calls the inputs "downscaling counterpart images". If downscaling means bicubic or average downsampling, each input pixel carries information from unmeasured locations, which real sparse measurements would not. Saying which one was used would settle this.

With the code or weights, the optimizer, the seeds, and the baseline settings added, a rerun of the experiments would be possible.

4. What specifically to fix

1. In Tables I and II, I could not reconcile the RMSE columns with the MSE columns. Using the mapping in Section IV.A.2, one grey level is 100/255 dB for RadioMapSeer, so the SRResNet MSE of 5.460 gives about 0.916 dB, but Table I reports 1.063 dB. For CKMImageNet, one grey level is 200/255 dB, so the SRResNet MSE of 76.64 gives about 6.866 dB, while Table II reports 7.684 dB. The reported dB values are higher than the MSE implies in every row of both tables, by roughly 10% to 30%. Maybe RMSE is computed on a different pixel set, for example with building pixels excluded. I would add the exact formula so readers can reconcile the two columns.

2. For the AoA results in Section IV.B.3 and Fig. 7, it would help to add a numeric error, for example, the mean absolute angle error in degrees, at each SR factor, next to the path loss numbers. I would also explain how the MSE loss handles the wrap-around between -180 and 180 degrees. The linear mapping puts these two equal angles at opposite ends of the grey scale, so the loss will treat them as far apart. A second small question is how the building pixels at -200 degrees fit a mapping defined for -180 to 180 (Section IV.A.2).
3. In Section IV.B.1, I would add at least one learned CKM baseline run on the same sparse input. RadioUNet [9] or Kriging [6] would fit well, since both are already cited in the Introduction.
4. In Section IV.B and Fig. 5, I would train each network with several seeds and report the mean and the spread. It would also help to report the spread across the 1,000 test maps. Right now, the statement that SRResNet is "more stable" than SRGAN (Section IV.B.2) is not supported by a number. Error bars in Fig. 5 would make this point easy to see.
5. Please say whether the low-resolution input is pure sampling, as in Eq. (1), or a downsampled image as described in Section III.B. If it is downsampled, one sentence on how this compares with real sparse measurements would help. A small wording point: H is called a diagonal matrix, but its size is $w'h' \times wh$, so "selection matrix" would be more precise.
6. For CKMImageNet in Section IV.A.1, I would state whether the 1,000 test maps share base stations or environments with the 11,064 training maps, as you did for RadioMapSeer.
7. In the Abstract, I would quote the CKMImageNet result, 7.684 dB at $4\times$ SR (Table II), next to the headline figure. I would also call bicubic interpolation just "interpolation" rather than "linear interpolation" in the Introduction and in Section IV.C.

5. Recommendation

Major revisions. I found this work honest, clearly written, and useful, and the label does not mean a low rating; it only reflects that some fixes need new runs. The idea holds: sparse values only, no building map, and SRResNet ranks first on both datasets (Tables I and II). I would add a numeric AoA metric (Section IV.B.3), one learned CKM baseline on the same sparse input (Section IV.B.1), and spread over seeds or test maps (Section IV.B.2). The rest needs only text: the RMSE formula that reconciles the MSE(pixel) and RMSE(dB) columns, the input construction in Eq. (1) vs. Section III.B, the CKMImageNet split, and the abstract.

Declarations

Potential competing interests: No potential competing interests to declare.