Radio Map Estimation with Deep Dual Path Autoencoders and Skip Connection Learning

William Locke, Nikita Lokhmachev, Yan Huang
Computer Science and Engineering
University of North Texas
Denton, Texas
{WilliamLocke, NikitaLokhmachev}@my.unt.edu, Yan.Huang@unt.edu

Xinrong Li
Electrical Engineering
University of North Texas
Denton, Texas
Xinrong.Li@unt.edu

***Abstract*—Radio Map Estimation (RME) is the task of predicting radio power at all locations in a two-dimensional area and at all frequencies in a given band. Accurate estimation is important for applications such as fingerprint based localization, user-cell association and power control in Massive Multiple Input Multiple Output (MIMO) systems, and path planning for Unmanned Aerial Vehicles (UAVs). The problem becomes especially challenging when transmitter locations are unknown and there are obstacles in the environment. In this case the predictions must be based on sampled radio power measurements and environmental information. Recent studies have shown that this can be accomplished using deep learning with Convolutional Neural Network (CNN) based autoencoders. However, these models mix environmental and signal information indiscriminately and do not share information between encoder and decoder. We propose the Dual Path Autoencoder to separate environmental and signal information and Skip Connection Autoencoders to increase information flow between encoder and decoder. We compare and contrast their performance with respect to sampling rate and learning architecture size. We find through experimentation that these two mechanisms independently improve predication accuracy over current state-of-the-art methods but do not appear to work in conjunction with each other.**

***Index Terms*—radio map estimation, deep learning, autoencoder, dual path, skip connection**

I. Introduction

RME is a crucial process in wireless communication networks of predicting radio power across a geographic area. This is useful in applications such as fingerprint based localization [1], user-cell association and power control in Massive MIMO systems [2], [3], and path planning for UAVs [4]. RME is conventionally performed utilizing interpolation [5] or modeling approaches [6]. Recently deep learning techniques have also been applied to the problem with great success [7]–[10].

Interpolation methods take measurements of radio power at specific locations and predict the power at points between and beyond them using mathematical and statistical methods, often with little or no information about the transmitters or physical environment. Modeling methods, on the other hand, proceed from specific information about the transmitters and physical environment and predict the power at any location by modeling the physical interactions between the radio signal and its environment. These methods are often more accurate but more computationally intensive than interpolation.

Deep learning methods learn from previously seen examples of radio wave propagation and generalize them to new and unseen contexts. Two recent studies have demonstrated the effectiveness of deep learning in contexts where interpolation or modeling would traditionally be used. In cases where transmitter location and environment are known, deep learning models predict radio power similarly to modeling methods, but much faster and using less computational resources [9], [11]. In cases where transmitter location is unknown, deep learning models use sampled measurements to predict the full radio map in a manner similar to interpolation, but with much higher accuracy and while also modeling signal-environment interactions [10], [12].

In this paper, we improve upon current state-of-the-art in RME with unknown transmitter locations [10] and make the following contributions. First, we propose a novel Dual Path Autoencoder architecture for processing signal and environmental information separately; all previous studies mix these inputs indiscriminately, despite the fact that they carry very different information with different encoding schemes. Second, we use UNet style skip connections [13] to increase RME accuracy by carrying information from encoder to decoder. The models in [9] also use UNet style architectures, but only in the context of known transmitter locations. Through experimentation, we show how each of these mechanisms independently improves the accuracy of predicted radio maps given different amounts of sampled data as input, but also how they do not appear to work in concert. We highlight several models that significantly improve upon the current state-of-the-art and compare their performance with respect to learning architecture sizes. All code is made publicly available.[1]

In Section II we define the problem of radio map estimation with unknown transmitter locations; Section III

[1]https://github.com/nikitalokhmachev-ai/radio-map-estimation-public/

describes our approach, including architecture schematics of our Dual Path Autoencoders and Skip Connection Autoencoders; Section IV gives our experimental design and results; and Section V discusses those results and directions for further research.

II. Problem Definition

Consider a two-dimensional area $\chi \subset \mathbb{R}^2$ of height $\mathcal{H}$ and width $\mathcal{W}$ with r sensors each located at a unique coordinate $\boldsymbol{x} \in \chi$. These r sensors measure the power in a given frequency band $\mathcal{F}$ at their coordinates $\boldsymbol{x}$, where $P(\boldsymbol{x}, f)$ denotes the power in frequency f at coordinate $\boldsymbol{x}$ (we assume effects of fast-fading are averaged out over small changes in the time or space dimension during measurement). The task of RME is to use the power measurements from these r sensors to infer the power $P(\boldsymbol{x}, f)$ at all points $\boldsymbol{x} \in \chi$ and all frequencies $f \in \mathcal{F}$.

To make the above problem statement suitable for machine learning, we first discretize χ into a $N_x \times N_y$ grid so that $\boldsymbol{x}$ is an index (x, y) where $x \in \{0, 1, \ldots, N_x - 1\}, y \in \{0, 1, \ldots, N_y - 1\}$ and each index $\boldsymbol{x}$ covers a rectangular area of size $\frac{\mathcal{W}}{N_x} \times \frac{\mathcal{H}}{N_y}$. We similarly discretize $\mathcal{F}$ by sampling a finite subset of frequencies and indexing them $f \in \{0, 1, \ldots, N_f - 1\}$, with N_f being the total number of frequencies sampled from the band $\mathcal{F}$. $P(\boldsymbol{x}, f)$ then denotes the power at the frequency indexed by f and averaged over the area of $\boldsymbol{x}$. To simplify our study at this early stage, in the following experiments we restrict f to a single frequency such that we only consider samples taken across the spatial dimension $\boldsymbol{x}$ with power $P(\boldsymbol{x})$.

To derive results that are useful in practice, we also consider environmental obstacles, in this case buildings. Buildings can be located at any position $\boldsymbol{x}$, often covering multiple adjacent positions. We consider these positions to have no corresponding radio power $P(\boldsymbol{x})$, however they do affect radio power in surrounding areas by reflecting radio waves back into the environment and by blocking or significantly attenuating radio waves passing through them. To allow the model to learn these interactions, we create an environment mask, an additional N_x by N_y grid that encodes every position $\boldsymbol{x}$ with either a 1, 0, or -1 depending on if it is the location of a sampled measurement, unsampled free space, or a building respectively. This is the same approach as taken in [10]; however, in Section III-B1 we propose a novel way of combining this environment mask with the sampled radio signal map, which we call the Dual Path Autoencoder.

III. Proposed Approach

A. Background

1) Convolutional Neural Networks: Convolutional Neural Networks (CNNs) are neural networks designed especially for image processing [14]. They operate by convolving the input with trainable filters, where a filter is an m by n array of learnable weights $\mathbf{w}$ that moves over the input $\mathbf{x}$ and outputs values $\mathbf{o}$ according to Eq. (1). The outputs of these convolutions are called "feature maps" and capture patterns within the image such as edges and shapes. Multiple filters are passed over an image at each layer, creating multiple feature maps. These feature maps are then passed through non-linear activations and pooling before being passed to later layers where the process is repeated. This allows the model to extract more abstract features from features it has already detected at earlier layers.

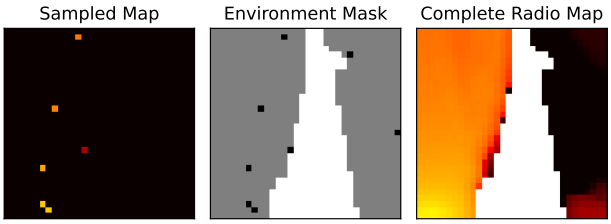


Fig. 1: Example N_x by N_y grid, where $N_x = N_y = 32$. Sampled Measurements and Environment Mask are inputs to the model, which tries to reconstruct the Complete Radio Map.

$$\mathbf{o}_{i,j} = \Sigma_{k=0}^{m-1} \Sigma_{l=0}^{n-1} \mathbf{w}_{k,l} \mathbf{x}_{i+k,j+l} \tag{1}$$

The input to a CNN can either have one channel, as in grayscale images, or multiple channels, as in color images. Note that Eq. (1) doesn't change in the latter case; the filters passing over an input with three RGB channels will also have three channels, and these filters will convolve all three channels together. This makes sense for images, where the different channels are all of a similar type and encoded in the same way. However, we show through experimentation that it is suboptimal for RME, where the different channels carry fundamentally different kinds of information that, after convolution, cannot be separated again for use in subsequent layers. We address this limitation in Section III-B1 with Dual Path Autoencoders.

2) Deep Learning Autoencoders: Autoencoders are a family of deep learning architectures that are trained to take an input, reduce it to a lower-dimensional representation (often referred to as a "latent space representation"), and then restore it to its original dimensionality in as faithful a recreation of the input as possible [15]. It is composed of two parts, the encoder and decoder, which compress and decompress the input information respectively. If we take χ to be the original input, λ to be its latent space representation, and ϵ and δ to be the encoder and decoder respectively, then we can state $\lambda = \epsilon(\chi)$ and $\chi = \delta(\lambda)$, so the complete autoencoder carries out the function $\chi = \delta(\epsilon(\chi))$.

Deep Learning Autoencoders using CNNs as encoder and decoder have made tremendous gains in image generation, such as in OpenAI's DALL-E [16]. A limitation of traditional autoencoder architecture is that information

passing between the encoder and decoder can only go through a single low-dimensional bottleneck. This is intentional so as to force the model to learn generalizable latent space representations. However, for reconstructing full radio maps from sampled maps and environment masks, this bottleneck may lose too much information and prevent information from flowing effectively from encoder to decoder (and vice-versa). We address this limitation using skip connections in Section III-B2.

B. Our Approaches

1) Dual Path Autoencoders: To address the fact that the radio map and environment mask encode different types of information that could be useful to different layers of the encoder or decoder, we propose a Dual Path Autoencoder that can pass either the sampled radio map or the environment mask separately from the convolved feature maps. This enables us to keep a clean copy of either of the two input channels (or, in one experiment, both) that we can attach to the inputs of later layers, giving them direct access to sampled or environmental information that would otherwise be subsumed in previous convolutions. An illustration of the Dual Path Autoencoder architecture is shown in Fig. 2.

Our hypothesis is that the different information encoded in the two channels will offer different benefits to either encoder or decoder. We hypothesize that the encoder should benefit most significantly from the environment mask, which contains building and sample locations, in order to create a more accurate encoding of the environment for radio wave propagation that it will then pass to the decoder. The decoder in turn should benefit most from the sampled map, which contains actual measurements of radio power, since its job is to reconstruct a full signal map that agrees with these measurements. We believe the Dual Path Autoencoder architecture could be applied in other contexts where the accurate encoding and decoding of one variable depends on its interaction with other variables of different types.

2) Skip Connection Autoencoders: Skip connections were introduced in [17] and [13] and are connections

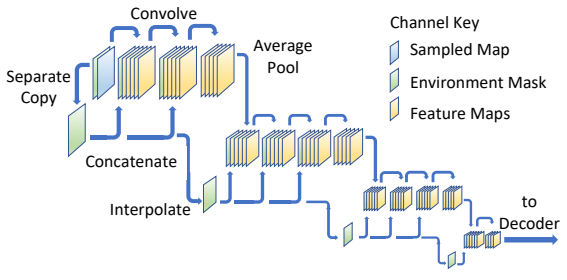


Fig. 2: Encoder side of a Dual Path Autoencoder, passing the environment mask separately. The decoder is the mirror image of the encoder, with either the environment mask or sampled map passed to it (see Fig. 3).

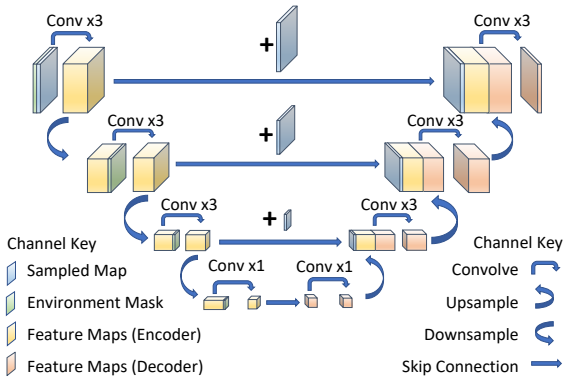


Fig. 3: Skip Connection Autoencoder with UNet style skip connections and Dual Path Encoder and Decoder. For visual clarity, we have combined the three convolution layers at each resolution into a single block and omitted the separate channel path shown in Fig. 2.

between inputs and outputs that go across one or more convolutions. They allow information to pass more efficiently between input and output and often improve performance and expedite training. ResNet style skip connections [17] skip one or two convolutions and can be applied to most CNN architectures, while UNet style skip connections [13] skip between encoder and decoder and are thus specific to autoencoder architectures. We experiment with both in the models below, but following [9] we focus mainly on UNet style skip connections.

UNet style skip connections pass convolved feature maps from encoder to decoder at specific points in the architecture. These transfers are illustrated in Fig. 3. Passing the convolved feature maps offers two major advantages. First, going from encoder to decoder, the feature maps from the encoder carry detailed information that is closer to the source input. As such, they relax the bottleneck at the latent space representation and assist the decoder in making more accurate predictions. Second, going from decoder to encoder, the back-propagation of error can also travel along these skip connections, allowing the encoder to learn and update its filters much more directly from the output of the model and therefore create new feature maps that better assist the decoder in reconstructing the complete radio maps.

IV. Experimental Design and Results

A. Dataset

To train and test these models, we create a dataset of simulated radio maps using the generators provided in [10].[2] Specifically, we use their Insite Map Generator, which generates roughly $90m \times 90m$ radio maps and environment masks drawn from a $700m \times 700m$ section of downtown Rosslyn, Virginia. For each map, we simulate radio propagation by setting (virtually) two

[2]https://github.com/fachu000/deep-autoencoders-cartography

transmitters at random locations in the $700m \times 700m$ area and propagating signals at 1400 MHz with a 5 MHz bandwidth. We then take a $90m \times 90m$ square from this larger area as our radio map and environment mask. The received radio power in dBW at all locations is calculated using a modeling method as discussed above in Section I, specifically the *shooting and bouncing ray method* [6]. As suggested in [9], we truncate the radio power at a lower bound of -134 dBW and then scale all values between 0 and 1. Radio and environment maps are discretized as 32×32 pixel grids, with each pixel covering approximately $8m^2$.

As input to our model, we sample the radio power measurements at random locations on each map and set power at all other locations (including buildings) equal to zero. Our train dataset consists of 225,000 radio maps sampled at rates drawn uniformly between 1% and 40% of total pixels (between 10 and 410 pixels). Our validation and test datasets consist of 24,990 and 25,000 radio maps respectively that are sampled at set intervals between 1% and 40% to allow us to compare model performance at specific sampling rates.[3] Each map is generated independently and has two channels: sampled radio map and environment mask. The latter encodes sampling locations, building locations, and unsampled free space as 1, -1, and 0 respectively. See Fig 1 on page 2 for an illustration of both channels of the input.

B. Models

In the experiments below we compare 12 different model architectures: one baseline and 11 new architectures. The baseline is our reimplementation of the state-of-the-art autoencoder in [10], trained on our dataset described in Section IV-A above. Our new architectures are split into two groups: Dual Path Autoencoders as described in Section III-B1 and Skip Connection Autoencoders as described in Section III-B2.

[3]The test set sampling rates run from 2% to 40% at intervals of 2%.

The Dual Path Autoencoders separate either the sampled map or the environment mask to pass along separately through the encoder and decoder. We experiment with the following combinations:

- Dual$_{map}$: map to encoder and decoder
- Dual$_{mask}$: mask to encoder and decoder
- Dual$_{map\text{-}mask}$: map to encoder, mask to decoder
- Dual$_{mask\text{-}map}$: mask to encoder, map to decoder
- Dual$_{input}$: map and mask to encoder and decoder

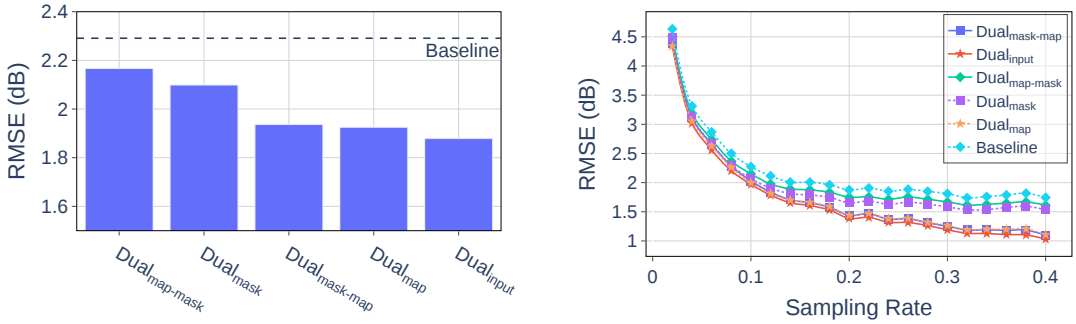
The Skip Connection Autoencoders combine UNet style skip connections with either the baseline architecture or our Dual Path Autoencoders. We experiment with the following combinations.

- Skip: skip connections with baseline
- Skip$_{residual}$: ResNet and UNet skip connections with baseline
- Skip$_{map}$, Skip$_{mask}$, Skip$_{map\text{-}mask}$, Skip$_{mask\text{-}map}$, Skip$_{input}$: Skip Connections with corresponding Dual Path architectures

C. Results

1) Dual Path Autoencoder: We find that the Dual Path architecture improves upon the performance of the state-of-the-art autoencoder, and that it is sensitive especially to which channel is passed to the decoder. The models that pass the sampled map to the decoder (Dual$_{map}$, Dual$_{mask\text{-}map}$, and Dual$_{input}$) see the largest gains in performance, while the models that pass only the environment mask to the decoder (Dual$_{mask}$ and Dual$_{map\text{-}mask}$) also improve over the baseline but see smaller gains. Dual$_{input}$ achieves the highest performance of all the Dual Path models.

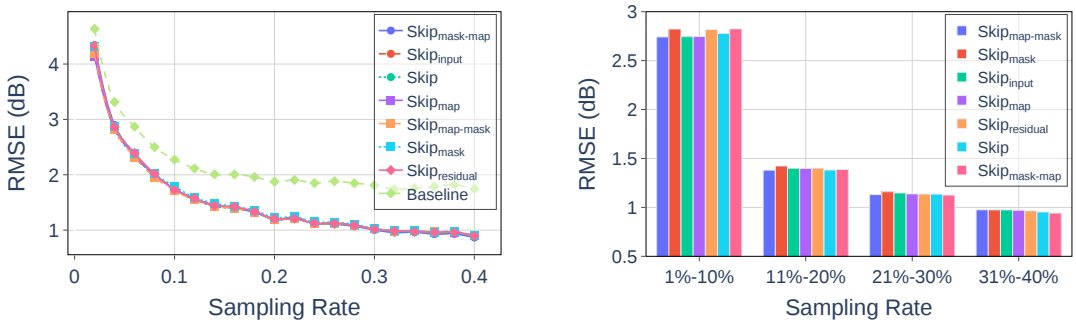
Model performance is also sensitive to the number of sampled measurements. Below 10% all models perform similarly, but above 10% they start to separate into two distinct groups based on which channel is passed to the decoder, as described above. Most likely this is because the sampled map becomes more informative at higher sampling rates. We had hypothesized that the



(a) RMSE across entire test dataset.

(b) RMSE broken down by sampling rate.

Fig. 4: RMSE of Dual Path Autoencoders



(a) RMSE broken down by sampling rate.

(b) RMSE binned by sampling rate.

Fig. 5: RMSE of Skip Connection Autoencoders

environment mask would be similarly informative for the Encoder, but looking at the Results this does not seem as strongly supported; there is no clear separation between models that pass the environment mask to the Encoder and those that do not. The performance of models that pass the sampled map to the decoder draws close to Skip Connection models at higher sampling rates (Fig. 6).

2) Skip Connection Autoencoders: We run a similar set of experiments on our Skip Connection Autoencoders, looking at RMSE across the entire dataset and at specific sampling rates.

We observe that all Skip Connection Autoencoders significantly outperform the baseline, and all perform very similarly to each other (Fig. 5a). Significantly, there is no clear separation between Skip Connection Autoencoders that also employ Dual Path architecture and those that do not, despite the improvement that Dual Path architecture shows over the baseline when applied by itself.

From Fig. 5a, it appears that the Skip Connection models all improve with increased samples at very close to the same rate. However, binning the sampling rates and taking RMSE across bins (see Fig. 5b) reveals that relative model performance is sensitive to sampling rate, and the models that perform best at higher sampling rates are not those that perform best at lower sampling rates.

3) Comparing Dual Path and Skip Connections: Comparing RMSE across the entire dataset, Skip Connection Autoencoders outperform Dual Path Autoencoders at all sampling rates. However, at higher sampling rates, the top performing Dual Path Autoencoders (those that pass the sampled map to the decoder) perform much more similarly to Skip Connection Autoencoders than to either the bottom performing Dual Path Autoencoders or the baseline (Fig. 6).

The different architectures also differ significantly in size (Fig. 7). Most of the Dual Path Autoencoders only add about 4,000 trainable parameters to the baseline model while improving performance significantly. Meanwhile Skip Connection Autoencoders add on the order of 20,000 trainable parameters to the model, while

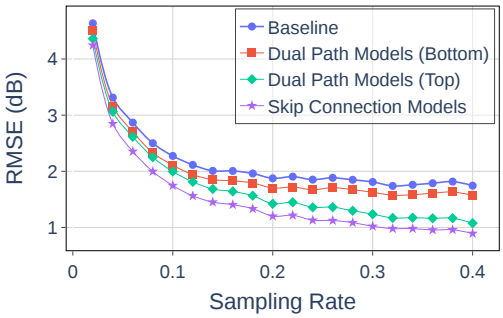


Fig. 6: Average RMSE of bottom performing Dual Path models, top performing Dual Path models, and Skip Connection models.

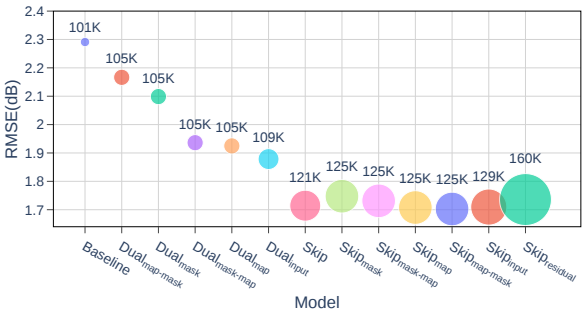


Fig. 7: Models ordered by size, diameter and number show number of trainable parameters, height on Y axis shows RMSE.

$\text{Skip}_{residual}$ adds an additional 40,000 above that. For practical applications, it is important to take into account the amount of sampled data available and the desired size of the model when selecting an architecture.

V. Conclusion and Future Work

We have proposed two architectures that improve upon current state-of-the-art for RME, Dual Path Autoencoders and Skip Connection Autoencoders. Motivation for these architectures stems from the demands of RME and the limitations of the current state-of-the-art models.

Dual Path Autoencoders address the fact that sampled radio maps and environment masks encode different types of information, but in current state-of-the-art models are mixed together indiscriminately and are no longer separable at deeper layers of the model. The fact that Dual Path Autoencoders improve upon the baseline most drastically at higher sampling rates when passing the sampled map to the Decoder suggests that the greatest gain from these models could come from their ability to pass signal information directly from the input to the decoder, a trait they share in common with Skip Connection Autoencoders. Dual Path Autoencoders are also significantly smaller than Skip Connection Autoencoders, usually by about 16,000 parameters.

Skip Connection Autoencoders address the limitation that information is lost between encoder and decoder in current state-of-the-art models. Skip connections in RME allow the decoder more direct access to the original inputs when reconstructing radio maps, and they allow the encoder more direct access to reconstruction error when updating its parameters. The dominance of all Skip Connection models over every other model architecture explored in this paper demonstrates their effectiveness as a learning tool in RME.

An outstanding question is why these two architecture designs don't seem to work in concert to improve performance beyond their individual application. One possibility is that Skip Connection Autoencoders and the top performing Dual Path Autoencoders might use the same strategy, i.e. passing signal information directly from the top layers of the encoder to those of the decoder with very little intervening convolution or interpolation. Dual Path Autoencoders do this explicitly by passing the unconvolved signal map to the decoder, while Skip Connection Autoencoders could learn a similar strategy by dedicating some filters to copying the signal map and then passing the resulting feature maps to the decoder. In this case, combining the two might simply provide redundant channels of information that increase the risk of overfitting. However, this questions the validity of the assumption that the sampled map and environment mask carry unique information that is lost when they are convolved together, an assumption that appears to hold true when comparing baseline autoencoders to Dual Path Autoencoders without skip connections. Therefore further investigation into how these mechanisms interact and facilitate RME still has potential for discovery and improvement.

ACKNOWLEDGMENT

Research was sponsored by the Army Research Laboratory and was accomplished under Cooperative Agreement Number W911NF-23-2-0014. The views and conclusions contained in this document are those of the authors and should not be interpreted as representing the official policies, either expressed or implied, of the Army Research Laboratory or the U.S. Government. The U.S. Government is authorized to reproduce and distribute reprints for Government purposes notwithstanding any copyright notation herein.

REFERENCES

[1] Yapar, R. Levie, G. Kutyniok, and G. Caire, "Locunet: Fast urban positioning using radio maps and deep learning," in *ICASSP 2022 - 2022 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 4063–4067, 2022.

[2] D. Bethanabhotla, O. Y. Bursalioglu, H. C. Papadopoulos, and G. Caire, "Optimal user-cell association for massive mimo wireless networks," *IEEE Transactions on Wireless Communications*, vol. 15, no. 3, pp. 1835–1850, 2016.

[3] T. Van Chien, T. Nguyen Canh, E. Björnson, and E. G. Larsson, "Power control in cellular massive mimo with varying user activity: A deep learning solution," *IEEE Transactions on Wireless Communications*, vol. 19, no. 9, pp. 5732–5748, 2020.

[4] S. Zhang and R. Zhang, "Radio map based path planning for cellular-connected uav," in *2019 IEEE Global Communications Conference (GLOBECOM)*, pp. 1–6, 2019.

[5] A. B. H. Alaya-Feki, S. B. Jemaa, B. Sayrac, P. Houze, and E. Moulines, "Informed spectrum usage in cognitive radio networks: Interference cartography," in *2008 IEEE 19th International Symposium on Personal, Indoor and Mobile Radio Communications*, pp. 1–5, 2008.

[6] H. Ling, R.-C. Chou, and S.-W. Lee, "Shooting and bouncing rays: calculating the rcs of an arbitrarily shaped cavity," *IEEE Transactions on Antennas and Propagation*, vol. 37, no. 2, pp. 194–205, 1989.

[7] X. Han, L. Xue, F. Shao, and Y. Xu, "A power spectrum maps estimation algorithm based on generative adversarial networks for underlay cognitive radio networks," *Sensors*, vol. 20, no. 1, 2020.

[8] X. Han, L. Xue, Y. Xu, and Z. Liu, "A radio environment maps estimation algorithm based on the pixel regression framework for underlay cognitive radio networks using incomplete training data," *Sensors*, vol. 20, no. 8, 2020.

[9] R. Levie, Yapar, G. Kutyniok, and G. Caire, "Radiounet: Fast radio map estimation with convolutional neural networks," *IEEE Transactions on Wireless Communications*, vol. 20, no. 6, pp. 4001–4015, 2021.

[10] Y. Teganya and D. Romero, "Deep completion autoencoders for radio map estimation," *IEEE Transactions on Wireless Communications*, vol. 21, no. 3, pp. 1710–1724, 2022.

[11] R. Levie, Yapar, G. Kutyniok, and G. Caire, "Pathloss prediction using deep learning with applications to cellular optimization and efficient d2d link scheduling," in *ICASSP 2020 - 2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 8678–8682, 2020.

[12] Y. Teganya and D. Romero, "Data-driven spectrum cartography via deep completion autoencoders," in *ICC 2020 - 2020 IEEE International Conference on Communications (ICC)*, pp. 1–7, 2020.

[13] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015* (N. Navab, J. Hornegger, W. M. Wells, and A. F. Frangi, eds.), (Cham), pp. 234–241, Springer International Publishing, 2015.

[14] K. O'Shea and R. Nash, "An introduction to convolutional neural networks," *CoRR*, vol. abs/1511.08458, 2015.

[15] G. E. Hinton and R. R. Salakhutdinov, "Reducing the dimensionality of data with neural networks," *Science*, vol. 313, no. 5786, pp. 504–507, 2006.

[16] A. Ramesh, M. Pavlov, G. Goh, S. Gray, C. Voss, A. Radford, M. Chen, and I. Sutskever, "Zero-shot text-to-image generation," *CoRR*, vol. abs/2102.12092, 2021.

[17] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 770–778, 2016.