Deep Learning-Driven Approaches for Uncertainty-Aware WiFi Indoor Localization

Zhaoguang Yi

Master of Science

School of Informatics

University of Edinburgh

2023

Abstract

This dissertation introduces two uncertainty-aware WiFi localization methodologies based on deep learning for RSSI RadioMap, exemplifying high reliability. In essence: 1) We leverage the Transformer model to semantically model AP Fingerprints for localization prediction. Concurrently, Evidential Deep Regression facilitates the model's self-supervised learning for uncertainty estimation. 2) We employ a Transformer-based autoencoder to unsupervisedly learn robust fingerprint embedding with spatial position associations. This embedding's similarity is utilized for WKNN-based localization. The methods were thoroughly validated in three campus environments. They are presented through both visualization and quantitative evaluations, showcasing exceptional localization accuracy and highly correlated uncertainty estimation. Remarkably, our approaches surpass the state-of-the-art proprietary industrial WiFi RSSI localization algorithms' results Huawei.

Our contributions are deemed highly innovative. Our research group is in the process of submitting our findings to premier academic conferences such as IEEE-ICRA or venues of equal distinction. For a succinct overview, readers can peruse the concise **internal preprint draft**, which provides a good snapshot of this 40-page dissertation.

Declaration

I declare that this thesis was composed by myself, that the work contained herein is my own except where explicitly stated otherwise in the text, and that this work has not been submitted for any other degree or professional qualification except as specified.

(*Zhaoguang Yi*)

Acknowledgements

I would like to extend my heartfelt gratitude to Chris Xiaoxuan Lu, Qiyue Xia, Peize Li, and Xiangyu Wen from the University of Edinburgh. Without their invaluable guidance and mentorship, the successful completion of this project would have been unattainable. I also wish to express my sincere thanks to Rory Hughes, Francisco Zampella, and Firas Alsehly from Huawei Edinburgh Research Center for their pivotal contributions in radio map generation and their expert insights.

Contents

Chapter 1

Introduction and Background

1.1 Wi-Fi Localizatation

Figure 1.1: RSSI-based Indoor Wi-Fi Localization [1]

Indoor positioning using WiFi represents a valuable methodology, primarily due to the inherent limitations of GPS satellite signals, which are often obstructed by buildings. The WiFi positioning technology can offset these impediments, providing a significant advantage. An additional benefit of WiFi-based positioning systems is that they do not necessitate the integration of extra hardware, thereby simplifying deployment procedures. Furthermore, this approach proves to be highly versatile, as evidenced by a myriad of practical applications that are outlined below.

Navigation: Wi-Fi indoor positioning can aid in improved navigation for individuals in large indoor spaces such as airports, shopping malls, and museums, by providing

accurate location information to help users easily find their way and locate specific points of interest. Additionally, robots can benefit from this technology as well [2].

Tracking: Wi-Fi indoor positioning can also be used to track the location of assets, such as equipment and inventory [3]. This can help businesses improve efficiency and reduce costs by quickly locating items that need to be moved or replenished. Moreover, this technology can improve safety and security by providing real-time tracking of people and assets, helping to identify potential hazards and respond promptly to emergencies.

Marketing and advertising: By leveraging the capability of tracking the position of users within a facility, businesses can deliver targeted marketing and advertising content that promotes specific products or services based on the user's location [4].

1.2 Project Motivtion

Indoor Localization services have been exponentially growing in importance and relevance over the past years. Among the various technologies available, WiFi-based positioning systems, particularly those based on Received Signal Strength Indicator (RSSI) fingerprints, have drawn considerable interest due to their low cost and high applicability. The WiFi positioning based on RSSI (Received Signal Strength Indicator) fingerprints possesses significant application potential in areas such as device tracking [3], indoor navigation, and advertising. RSSI fingerprint-based positioning typically consists of two main phases: the offline and online stages. In the offline stage, a "fingerprint" database is generated. This database contains vectors, each of which represents the RSSI value of each MAC address corresponding to their physical locations.

In the online stage, the positioning system utilizes the RSSI values collected by the device to be located, which queries the database that was created during the offline stage. An algorithm then compares the observed RSSI values to those in the database and returns the best-matching location. This two-step process enables WiFi-based positioning systems to offer an effective solution for indoor positioning.

However, creating an RSSI fingerprint database through the offline stage could be a challenging task, collecting the RSSI values at various coordinate points within a specific environment is not only time-consuming but can also lead to inconsistencies due to human error and variability in signal strength over time. In industry, crowd-sourcing has emerged as a popular method for data acquisition. Crowdsourcing approaches typically perform a global optimization using data from WiFi and IMUs, our dataset

originates from a top-tier, proprietary algorithm in the industry. This approach involves collecting data from a wide array of volunteer devices, such as smartphones and tablets, which are often equipped with WiFi receivers and IMUs. All data used in this manner is collected with the volunteers' consent, and sensitive information is appropriately anonymized to ensure privacy. Harnessing the power of these ubiquitous devices allows for the passive collection of a substantial and diverse amount of data at a low cost.

Despite the numerous advantages of crowd-sourcing methods, they also introduce a significant level of data noise. The diverse range of devices used in data collection can lead to variances in the recorded RSSI values due to differences in the hardware characteristics [**nnoise**]. Furthermore, the uncontrolled conditions under which the data are gathered - considering the volunteers' arbitrary movements and the ever-changing environmental factors - contribute to the inconsistencies in the collected data.

These irregularities and noise in the data make it difficult to apply regression-based machine learning models directly. These models, although effective under controlled circumstances, may produce significant errors when applied to noisy and unpredictable data sets. The issue of overfitting becomes particularly prominent, as the models might attempt to fit the noise rather than the actual signal pattern. [5].

Accurately mapping fingerprints to locations can be challenging due to the inherent noise and non-linearity of Wi-Fi signals. Traditional methods, such as calculating a distance metric between the measurements collected during the training phase and those of the target fingerprints during testing [6], are commonly used. However, this approach can result in inaccuracies due to the variability in Wi-Fi signals caused by factors such as signal attenuation, multi-path propagation, and interference from other devices.

In response to the outlined scenario, we aim to develop a method that utilizes deep learning to reduce noise and discern internal patterns, thereby enhancing the existing WiFi positioning algorithms. By applying deep learning techniques, we seek to construct a robust model that can automatically learn and generalize the hidden patterns within the signal data, thereby facilitating greater accuracy in indoor localization.

This approaches leverages the power of deep learning, which has shown immense promise in dealing with complex and noisy data, for the improvement of WiFi positioning. It is anticipated that by integrating this approach, we can significantly diminish the detrimental impact of environmental noise, a common issue in WiFi-based systems, and enhance the capability to discern underlying patterns in signal propagation. This, in turn, has potential implications for the refinement of the existing WiFi positioning algorithms.

1.3 Thesis Structure

Chapter 2: Literate Review: Foundations of WKNN and Deep Learning-based Localization

This chapter elucidates the fundamental principles behind the Weighted K-Nearest Neighbors (WKNN) algorithm and delves into the integration of deep learning techniques for localization purposes. It aims to provide the reader with a foundational understanding of the traditional and modern methodologies employed in the realm of RSSI fingerprinting.

Chapter 3: Uncertainty-Aware Deeo Regression for WiFi Positioning

Chapter 3 expounds on a novel method that harnesses the Uncertainty-aware Fingerprints Transformer model, complemented by an evidential deep regression technique. We validate our approach through empirical assessments conducted across three distinct environments within the University of Edinburgh campus: the Bayes Center Floor3, Informatics Forum Floor 1, and the Main Library Floor1. The findings from these experiments are juxtaposed with results obtained from traditional WKNN and cutting-edge industrial algorithms, highlighting the merits of our method.

Chapter 4: Expanding WKNN for Uncertainty Estimation

In this chapter, we present an advanced adaptation of the WKNN approach for localization. We introduce a novel metric— the dispersion among candidate locations—as an indicator of the inherent uncertainty in WKNN's predictions. Furthermore, this chapter sheds light on how global uncertainty metrics can effectively guide the adaptive selection of the hyperparameter 'K'.

Chapter 5: Neural WKNN through Transformer-based Autoencoder Embeddings

Chapter 5 delves into the transformative potential of harnessing Transformer-based autoencoders to obtain robust embeddings for WKNN, subsequently evolving it into a Neural WKNN. Through a series of rigorous evaluations, we demonstrate that this approach culminates in significantly enhanced localization accuracy.

Chapter 2

Literate Review: Foundations of WKNN and Deep Learning-based Localization

2.1 KNNs: Traditional Positioning in RSSI Fingerprinting

Weighted K-Nearest Neighbors (WKNN) is a popular method employed in RSSI fingerprinting-based positioning systems [7]. The primary objective of WKNN is to enhance the positioning accuracy by assigning weights to the K nearest neighbors rather than treating them with equal importance, as in the traditional K-Nearest Neighbors (KNN) method.

In the offline phase, the radio map is constructed by collecting the RSSI readings from various Wi-Fi access points at different reference points. During the online phase, the RSSI readings from the user's device are compared with the radio map. The algorithm identifies the K reference points with the most similar RSSI patterns as the 'nearest neighbors'. However, unlike the KNN method, WKNN assigns different weights to these neighbors based on their similarity to the observed RSSI pattern, typically following an inverse relationship - the higher the similarity, the greater the weight.

This approach allows WKNN to incorporate the confidence or reliability associated with each reference point into the positioning process, thereby improving accuracy. However, WKNN is not immune to the influence of multipath propagation and RSSI variability, which can impact the correct identification of nearest neighbors and consequently the positioning accuracy.

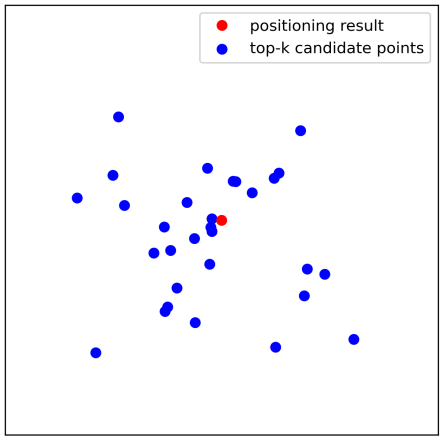


Figure 2.1: KNN for positioning

2.2 Deep Learning in RSSI Fingerprinting

In recent years, deep learning, a subfield of machine learning, has shown significant promise in improving the accuracy of RSSI fingerprinting-based positioning. Deep learning models, such as Convolutional Neural Networks (CNNs), Recurrent Neural Networks (RNNs), and their variants, have been employed to model the complex and nonlinear relationship between RSSI readings and physical locations [8].

CNNs, primarily used in image processing, have been adapted for RSSI fingerprinting by treating the RSSI maps from different access points as an image-like structure or matrix. In the offline phase, the CNN model is trained with these 'images' and their corresponding labels (locations). The convolutional layers of the CNN are adept at capturing local and global patterns in the RSSI readings, which contributes to a more robust positioning model.

RNNs, particularly Long Short-Term Memory (LSTM) networks, are effective in situations where temporal dynamics are important. For instance, they can be used in RSSI fingerprinting when the user is moving, and there is a temporal correlation between successive RSSI readings [9] [10]. The LSTM model is trained to recognize these temporal patterns and predict the current position based on the history of RSSI readings and movements.

Self-attention, an NLP-originated model structure, was also proved to be possible to model MAC addresses with RSSI. The paper "DRVAT: Exploring RSSI series representation and self-attention for WiFi indoor localization" [11] proposes a deep learning localization system, termed DRVAT, which is based on the distributed representation

vector (DRV) and self-attention among the pairs of MAC-RSSI. First, DRVs are obtained which represent dense features in low dimensionality through pre-training on all MAC-RSSIs. Then, a self-attention mechanism is used to capture the dependencies between different MAC-RSSIs.

One significant aspect of this paper is the introduction of pre-training and fine-tuning techniques. The proposed system can effectively utilize large amounts of unlabeled location data with minimal labeling effort. The authors have employed a training technique similar to BERT's [12], which randomly masks a position and trains the network to predict the masked position. During fine-tuning, the decoder is replaced with a decoder that predicts the masked positions. Through unsupervised pre-training on a large amount of data, the model can accurately learn to encode MAC-RSSI features [11].

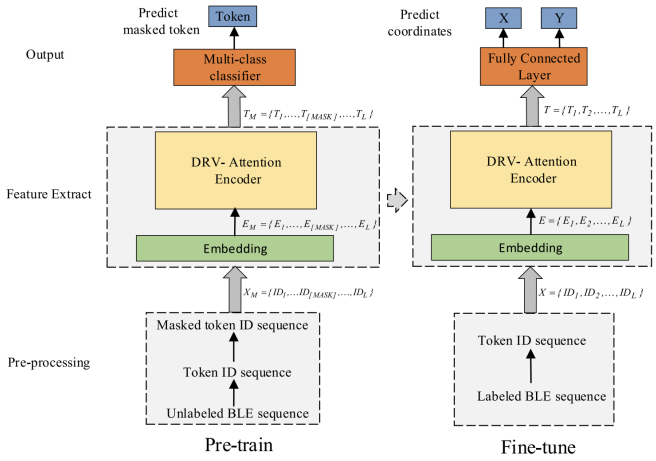


Figure 2.2: VDRVAT: Pre-training and Fine-tuning [11]

Deep learning models offer significant advantages in RSSI fingerprinting, such as the ability to automatically learn and extract features, handle high-dimensional data, and model complex nonlinear relationships. However, they come with their own set of challenges, such as the need for large training datasets, longer training times, and more computational resources. Furthermore, deep learning models are often considered as 'black boxes' due to their lack of interpretability, which can be a drawback in applications where understanding the model's decision-making process is crucial.

In the course of this project, another limitation of deep learning methods was identified by us, particularly evident using crowd-sourcing data. In data post-processing, a significant volume of data is processed through geographic location optimization algorithms to produce positioning results. These outcomes, however, were found to be

marked by inaccuracies and high levels of noise. If the deep learning model is misled by these noise-ridden results, its ability to generate more accurate outcomes is significantly compromised. Conversely, the Weighted K-Nearest Neighbors (WKNN) approach demonstrates resilience to this issue. By considering multiple candidates, WKNN is able to balance out the influences, yielding a result characterized by reduced noise.

2.3 Summary

This chapter provided an overview of widely adopted traditional positioning methods within the industry, with a particular focus on the Weighted K-Nearest Neighbors (WKNN) approach. In addition to these, a myriad of WiFi positioning methodologies underpinned by deep learning principles were discussed. In the ensuing chapter, an innovative technique is to be introduced, which improves the precision of traditional positioning methodologies through the application of nonlinear weights to each distinct RSSI value. This proposition introduces a nuanced method to address the limitations of conventional systems, paving the way for more accurate positioning outcomes.

Chapter 3

Uncertainty-aware Deep Regression for WiFi Positioning

3.1 Introduction

Estimation of uncertainty is a vital area in deep learning, as it allows us to gauge whether the outputs generated by a model are reliable or not. Evidential deep learning [13] is one of the most common methods used to this end. In the realm of Wi-Fi localization, it takes on a significant role. We have integrated this technique into our Wi-Fi positioning model, enabling it to assess the reliability of the received fingerprinting data (FP), and subsequently make predictions. This novel approach is aimed at enhancing the accuracy and robustness of Wi-Fi localization, paving the way for more sophisticated and dependable applications.

In this chapter, we introduce a novel network structure that maps each input Access Point (AP) to a lexical embedding. By doing so, the individual characteristics and attributes of each AP are captured in a more meaningful and condensed form. The network then employs the Transformer's self-attention mechanism to sift through these embeddings, focusing on and extracting the most relevant AP correlations. This innovative approach allows the system to understand and prioritize key relationships between various APs, providing a more nuanced and effective way to interpret Wi-Fi localization data. The introduction of lexical embeddings coupled with self-attention fosters a more complex understanding of the spatial relationships between APs, which is essential for improving localization accuracy and efficiency.

3.2 Methodology

3.2.1 Model Structure

To extract more meaningful features for distance comparison, we employ a denoising encoder. The sheer richness and volume of data embedded within the radio map underscore an opportunity to leverage the advanced methodologies of deep learning. This, in turn, facilitates the extraction of a more refined and sophisticated fingerprint representation. Harnessing the prowess of deep learning for feature extraction can unlock a deeper, high-level understanding of the data, potentially escalating the precision of WKNN localization.

A prime candidate for achieving this objective is the transformer [14], a stalwart model in the field of Natural Language Processing (NLP), lauded for its high parallelizability and capacity to expand model complexity with additional layers. Moreover, the concept of word embedding, commonly applied in NLP, provides an intuitive methodology for representing individual Access Points (APs), enabling our model to capture correlations among APs effectively.

Contrary to the standard sequence-to-sequence transformation in a transformer, our task entails the transformation of a sequence of Access Points (APs), each designated by its MAC address and received signal strength (RSSI), into a dense feature vector, denoted as the Fingerprint Feature. In this context, each AP can be thought of as analogous to a word in Natural Language. A learnable word embedding is associated with each AP, which adeptly encapsulates its unique characteristics within the context of localization. This approach not only yields a meaningful representation of individual APs but also facilitates the learning of patterns and correlations among different APs, a feature made possible by the self-attention mechanism.

To incorporate the RSSI values linked to each AP within the model, we leverage a gating mechanism as shown in Figure 3.2.1. The distinct features of each AP are ushered through a gate, the degree of information allowed to permeate being dictated by the associated RSSI values. This mechanism ensures suitable representation of the RSSI values, facilitating information flow through each gate that is proportionate to the intensity of the affiliated signal. Notably, when an AP's RSSI is substantial, a considerable amount of information is permitted to pass through the gate. Conversely, if the RSSI is minimal or fails to be detected, little to no information can traverse the gate, thereby limiting its impact on the system.

This impact ultimately cascades to the attention mechanism, as the diminishment

in input features results in a corresponding reduction in the value, key, and query parameters. This forms a type of prior knowledge for the network: small and missing RSSI values are closer to each other and exert less influence on the system, while larger RSSI values carry a more significant and definitive quantity of information.

The attention mechanism is defined as:

$$\text{Attention}(Q,K,V) = \text{softmax}\left(\frac{QK^T}{\sqrt{d_k}}\right)V$$

Where Q, K, and V are the queries, keys, and values, respectively, and are calculated as:

$$Q = W_q \cdot X_{gated}$$

$$K = W_k \cdot X_{gated}$$

$$V = W_v \cdot X_{gated}$$

In the above equations, X_gated represents the gated AP embeddings; W_Q, W_K, and W_V symbolize the weight matrices for Q, K, and V, respectively; and d_k stands for the dimensionality of the queries and keys. We introduce a modification by scaling the AP embedding X before calculating Q, K, and V:

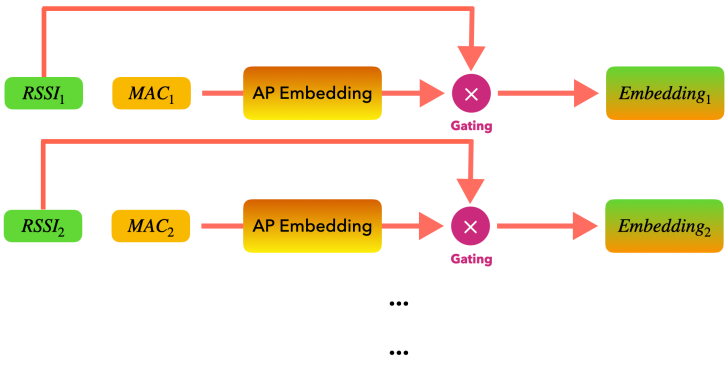


Figure 3.1: Gatting Embedding based on RSSI

The gating strategy introduced in this discourse serves to foster a more harmonious synchronization between the attention mechanism and the intrinsic properties of the Received Signal Strength Indicator (RSSI) signals encapsulated in the radio map. The driving rationale behind this alignment is the profound understanding that RSSI signals, with their unique characteristics, necessitate a tailored approach to ensure optimal information extraction.

To operationalize this strategy, we employ neural networks to learn a scaling coefficient tailored to the entirety of the RSSI spectrum. By doing so, we leverage the inherent capability of neural networks to adaptively discern patterns and relationships in data. The learned scaling coefficient then modulates the attention mechanism, ensuring that it aligns seamlessly with the nuances of the RSSI signals. This symbiotic interplay between the attention mechanism and the RSSI characteristics ensures enhanced localization precision, thus bolstering the overall efficacy of the proposed system.

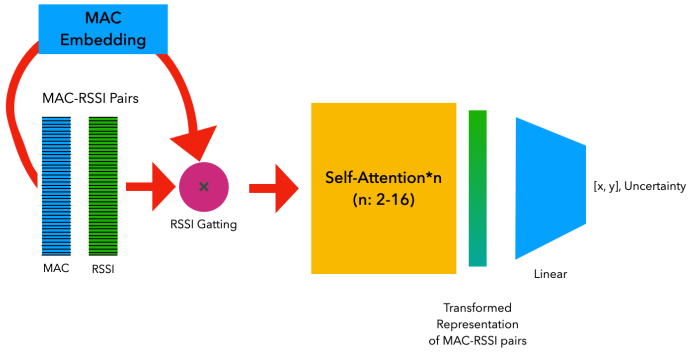


Figure 3.2: Uncertainty Regression Model Architecture

After a few self-attention layers, in the decoding process, we employ two distinct fully-connected neural networks that do not share parameters. One network is specifically tasked with predicting location, while the other is designated for the prediction of uncertainty. An additional uncertainty head is incorporated into the training scheme, facilitated by a tailored modification to the loss function. This innovative design enables the network to essentially assign confidence scores to its own predictions. By doing so, the network is able to strategically reduce the weight attributed to localization error through increased uncertainty. However, this advantage is counterbalanced by the introduction of an additional loss term, representing the cost associated with the uncertainty. This configuration creates a nuanced trade-off between prediction accuracy and the complexity introduced by the additional uncertainty component, providing a potentially rich avenue for further exploration and refinement in predictive modeling.

The detailed uncertainty estimation will be explained next section.

3.2.2 Uncertainty Estimation

In localization, it's crucial to recognize that noise can affect each input differently. In such a scenario, heteroscedastic aleatoric uncertainty needs to be considered [15], as it

encapsulates the data-dependent variations in the inputs. Since aleatoric uncertainty is inherently tied to the data, the loss function $L(\theta)$ must be modified to account for the uncertainty in the RSSI-based position estimate for each specific component.

In Evidential Deep Learning, the loss function is often designed not only to emphasize the predictive accuracy but also the model's uncertainty in its predictions. A commonly utilized loss function is given by:

$$L = 0.5 \times \text{MSE}(out, label) \times \exp(-s_d) + s_d$$

where $\text{MSE}(out, label)$ denotes the Mean Squared Error between the model's prediction and the actual label. The term s_d, representing the model's uncertainty or confidence about its predictions, is derived from the variance of the predictions (σ_d^2) through the relation:

$$s_d = \log(\sigma_d^2)$$

The logarithmic transformation ensures a positive and more stabilized representation of the model's uncertainty, making the optimization process more tractable.

As an engineering consideration, training for different noise environments may require adjusting the weights (A and B) of various components to achieve balance. This tailored approach ensures that the model is not only responsive to the unique characteristics of each environment but also maintains equilibrium, optimizing overall performance and accuracy:

$$L = A \times citerion(out, label) \times exp(-s_d) + B \times s_d$$

By explicitly modeling and quantifying uncertainty, our approach enhances the understanding of the inherent vagueness associated with wireless signal propagation and environmental factors. This not only leads to more robust predictions but also provides insightful information that could be instrumental in various applications where Wi-Fi localization is utilized.

In the subsequent section, we will present the experimental results of our uncertainty estimation and localization in a specific environment. This analysis includes a visualization that illustrates the complex interplay between the predicted locations and associated uncertainties.

3.3 Results and Discussion

Methods	Bayes F3			Forum F1			Main Libiray F1		
	Mean	Median	Max	Mean	Median	Max	Mean	Median	Max
Naive-WKNN	3.0009	2.45717	11.1487	3.6371	3.1127	15.5875	2.6733	2.1464	14.5017
Regression	2.2762	1.91042	14.4922	3.5603	2.8370	21.1554	2.65971	2.340926	12.0070
Huawei	2.3625	2.0358	12.9366	-	-	-	-	-	-

Table 3.1: Results on 3 Different Datasets

The table 3.1 presented showcases the results of different methods on three distinct datasets: Bayes F3, Forum F1, and Main Library F1. These results are categorized based on three performance metrics for each dataset: Mean, Median, and Maximum values.

The Naive-WKNN method consistently presents results across all three datasets. Its performance appears moderately stable, with mean values ranging from 2.6733 to 3.6371. The Regression method, intriguingly, outperforms Naive-WKNN in the Bayes F3 dataset when comparing the mean values but has a slightly larger maximum error. Huawei's method provides results only for the Bayes F3 dataset, rendering direct comprehensive comparison challenging. However, in this singular dataset, its mean performance situates between the Naive-WKNN and Regression methods.

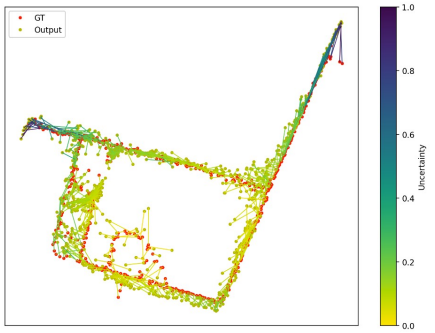
The Regression method manifests a notable precision, as particularly underscored in the Bayes F3 dataset where its average deviation outperforms that of Naive-WKNN. Nonetheless, the conspicuous max error observed in the Forum F1 dataset can potentially be mitigated through uncertainty estimation.

In Fig 3.3 the uncertainty results are visualized. To visualize the uncertainty-aware regression in 2D map, we depict the Ground Truth (GT) as red scatter points and the predictions as yellow scatter points. The connecting lines between them are color-coded to represent the visualized uncertainty. These visual representations are distinctly delineated for each of the three buildings. Furthermore, we have segmented the prediction errors and uncertainties into a grid format, where the depth of the grid color signifies frequency, providing a clear visual correlation between error magnitude and its associated uncertainty.

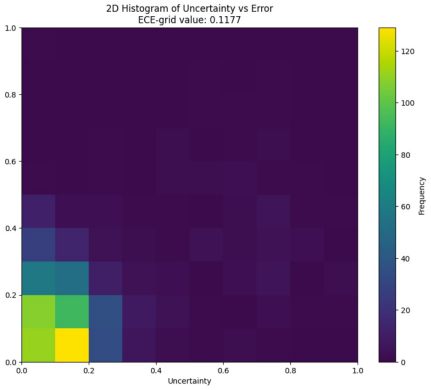
The visualized results strongly corroborate the efficacy of our model in yielding precise uncertainty predictions. In the map visualizations, it's conspicuously discernible that within the Main Library environment, predictions where our model indicated substantial errors were accompanied by heightened uncertainties. This pattern is not

just idiosyncratic to the Main Library but manifests in the other two environments as well, substantiated by intricate details that the visualizations elucidate.

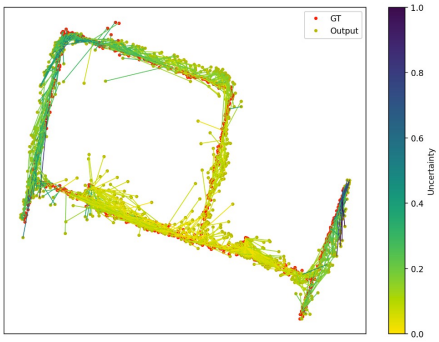
Complementing this, the binned and gridded Uncertainty-Error grid on the right accentuates the robust correlation between prediction errors and their associated uncertainties. Given the gridded nature of our results, we further employ the Expected Calibration Error (ECE) score to quantitatively encapsulate this relationship, offering a cohesive metric that reinforces the visual patterns observed.



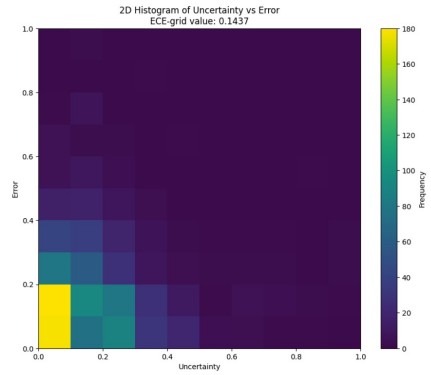
(a) Bayes F3 Prediction & GT



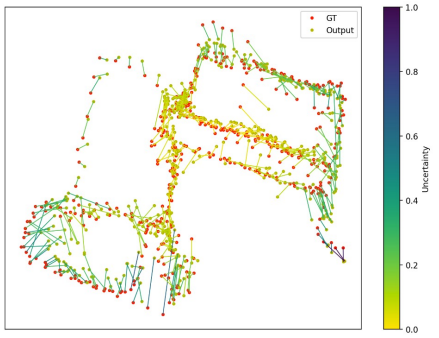
(b) Bayes F3 Uncertainty-error grid



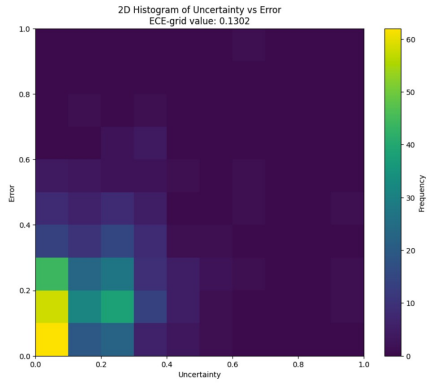
(c) Forum F1 Prediction & GT



(d) Forum F1 Uncertainty-error grid



(e) Main Library F1 Prediction & GT



(f) Main Library Uncertainty-error grid

Figure 3.3: Result Visualizations

3.4 Summary

In this chapter, we delved into the application of the deep evidential learning concept within the domain of Wi-Fi localization. A proficient machine learning model is not only gauged by its ability to produce accurate predictions but also by its capability to convey the confidence associated with those predictions. Such an introspective quality facilitates downstream fusion and enhances the robustness of integrated systems.

Empirical evidence underscores the model's adeptness at discerning unreliable portions within the Wi-Fi signal, marking them with associated uncertainties. This not only aids in achieving a holistic understanding of the Wi-Fi environment but also paves the way for reliable decision-making.

Furthermore, incorporating a degree of self-awareness in models, as seen with our uncertainty estimation, is paramount in real-world scenarios. It allows the system to gauge when to act upon a prediction, when to seek additional information, or when to resort to fallback mechanisms. Thus, the fusion of deep evidential learning with Wi-Fi localization transcends mere position estimation and serves as a testament to the evolution of intelligent systems that are both accurate and cognizant of their limitations.

Chapter 4

Uncertainty Estimation of WKNN and Adaptive 'K' Selection

4.1 Introduction

In the realm of machine learning applications, it is vital for a model to have an inherent capacity to gauge the uncertainty or confidence in its predictive outcomes [13]. This characteristic is of paramount importance in localization tasks, particularly since the errors inherent in a given system may propagate into subsequent processes. For instance, in a typical navigation scenario, Wi-Fi localization results are often integrated with Inertial Measurement Unit (IMU) data to perform sensor fusion. In cases where the Wi-Fi localization provides imprecise or uncertain results, the subsequent systems could be misled, leading to inaccurate overall positioning.

The traditional Weighted K-Nearest Neighbors (WKNN) algorithm, in its rudimentary form, does not offer a quantification of the uncertainty associated with its localization outcomes. However, our empirical observations suggest that the dispersion or standard deviation of candidate coordinates, obtained from the neighbour selection localization process, could serve as an effective proxy for quantifying positional uncertainty.

The estimation of uncertainty not only benefits subsequent systems by enabling them to filter out results with high positional errors, but also proves advantageous in optimizing certain hyperparameters within the WKNN algorithm itself. More specifically, we found that these uncertainty estimates could be effectively utilized in determining the optimal value for the critical hyperparameter 'K'. 'K' signifies the number of nearest neighbours considered in the WKNN algorithm, and its value has a direct bearing on

the performance of the localization system.

4.2 Methodology

4.2.1 WKNN Basics

The Weighted K-Nearest neighbours (WKNN) algorithm operates by identifying the 'k' nearest neighbours of a point based on a pre-defined criterion of proximity or 'closeness'. Conventionally, this proximity is quantified using the Euclidean or Manhattan distance metric. To refine this approach, the WKNN algorithm assigns more weight to the samples that more closely align with this specified criterion. This enhancement to the basic K-Nearest neighbours (KNN) methodology results in a more nuanced and focused strategy for neighbor selection, offering improved responsiveness to changes in proximity.

In the context of Wi-Fi fingerprinting localization, the application of a traditional WKNN algorithm, based on Received Signal Strength Indicator (RSSI), can be illustrated as follows:

We start by defining a set of RSSI fingerprint (FP) inputs, which serve as the observed signal strengths from various Access Points (APs). This set can be denoted as:

$$FP_t : \{RSSI_{1,t}, RSSI_{2,t}, RSSI_{3,t}, ..., RSSI_{i,t}, ..., RSSI_{N,t}\}$$

where 'i' represents the specific Access Point(AP) out of a total of 'N' APs, and 't' denotes the timestamp of the measurement.

Parallel to this, we have a reference set of RSSI values, which are the pre-recorded database at known locations from the same APs. This reference set, often referred to as the 'radio map'.

$$
\begin{aligned}
FP_1 &: \{RSSI_{1,1}, RSSI_{2,1}, ..., RSSI_{i,1}, ..., RSSI_{N,1}\} && Pos_1 \\
FP_2 &: \{RSSI_{1,2}, RSSI_{2,2}, ..., RSSI_{i,2}, ..., RSSI_{N,2}\} && Pos_2 \\
&\vdots && \vdots \\
FP_r &: \{RSSI_{1,r}, RSSI_{2,r}, ..., RSSI_{i,r}, ..., RSSI_{N,r}\} && Pos_r \\
&\vdots && \vdots \\
FP_R &: \{RSSI_{1,R}, RSSI_{2,R}, ..., RSSI_{i,R}, ..., RSSI_{N,R}\} && Pos_R
\end{aligned}
$$

where 'i' represents the specific Access Point (AP) out of a total of 'N' APs, and 'r' denotes the sample index in the database, which has a size of 'R'. Each fingerprint is paired with a corresponding position.

The difference in location fingerprint similarity between the t-th input signal and the r-th reference in the database can be defined as:

$$\Delta_{t,r} = criterion(FP_t, FP_r)$$

where $\Delta_{t,r}$ represents the difference or dissimilarity between the fingerprint of the t-th input signal and the fingerprint of the r-th reference in the database. The function criterion$(\cdot)$ captures the specific similarity measure employed to assess the distance between the two fingerprints.

To identify the closest neighbours, the delta values can be sorted in ascending order (or descending order if using a similarity measure), enabling the selection of the top-k nearest neighbours based on their proximity to the input signal.

Assuming we have already selected the positions corresponding to the k nearest neighbours, we can utilize the delta values as weights to compute a weighted mean of these positions.

$$Pos_t = \frac{[Pos_1, Pos_2, ...Pos_k] \cdot [\Delta_1, \Delta_2, ...\Delta_k]}{k \cdot \sum_i^k \Delta_i}$$

where Pos_t represents the estimated position of the input FP at timestep t. $[Pos_1, Pos_2, \ldots, Pos_k]$ represents the positions of the top-k nearest neighbours, and $[\Delta_1, \Delta_2, \ldots, \Delta_k]$ corresponds to the delta values (weights) associated with the nearest neighbours. The value k denotes the number of nearest neighbours considered. The denominator, $k \cdot \sum_i^k \Delta_i$, normalizes

the weights by the sum of all delta values multiplied by k.

This weighted mean calculation combines both the proximity of each neighbour (indicated by the delta value) and their respective positions, resulting in a more accurate estimation of the input FP's location.

Presented below is the pseudo-code for the Weighted K-Nearest Neighbors (WKNN) algorithm. Herein, K symbolizes the number of neighbors considered for computation, while a denotes a scaling hyperparameter, crucial for modulating the influence of RSSI feature vector in the prediction process. The integration of such hyperparameters not only provides flexibility in tuning the model for varied datasets but also ensures a refined balance between local adaptability and generalization.

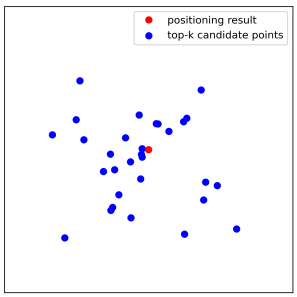
Algorithm 1 Improved WKNN for FP positioning

procedure WKNN($FP_t, FP_{\text{ref}}, k, a$)
    $FP_t \leftarrow a^{FP_t/20}$
    $FP_{\text{ref}} \leftarrow a^{FP_{\text{ref}}/20}$
    Initialize an empty list Δ
    for each FP_r in FP_{ref} **do**
        Compute $\Delta_{t,r}$ using $criterion(FP_t, FP_r)$ (either Cosine or Manhattan)
        Add $\Delta_{t,r}$ to Δ
    end for
    Sort Δ in ascending (or descending) order and select the top nearest k entries to get $\Delta_{\text{top\_k}}$
    Identify the positions $Pos_{\text{top\_k}}$ corresponding to $\Delta_{\text{top\_k}}$ in $FP_{\text{reference}}$
    Compute $Pos_t = \frac{[Pos_1, Pos_2, \ldots Pos_k] \cdot [\Delta_1, \Delta_2, \ldots \Delta_k]}{k \cdot \sum_{i=1}^{k} \Delta_i}$
    return Pos_t
end procedure

4.2.2 Estimating Uncertainty using Dispersion of Candidate Locations

The methodology adopted by the WKNN algorithm is to select 'K' nearest neighbours based on the input feature vectors. The inherent uncertainty associated with these estimates can be gauged by analyzing the dispersion of the selected candidate locations. The standard deviation of these locations serves as a practical measure of the dispersion, providing us with a quantifiable estimate of the uncertainty involved in the localization process.

In a scenario where the chosen candidate locations are significantly dispersed, we infer that the database does not possess a reliable matching position that corresponds precisely to the input features. This implies a high degree of uncertainty in our estimate, as shown in Fig.4.1. Conversely, if the dispersion or standard deviation among the candidate locations is relatively small, it implies a stronger match between the input features and the database, thus indicating a lower degree of uncertainty, as shown in Fig.4.2. By employing this method, we are able to quantify and interpret the degree of confidence associated with our localization estimates.



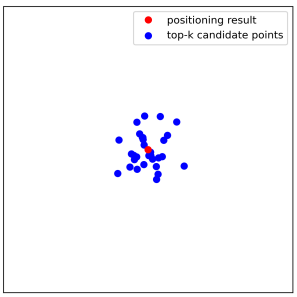


Figure 4.1: Estimated as High Uncertainty Figure 4.2: Estimated as Low Uncertainty

Through rigorous experimentation, we have observed a substantial correlation between uncertainty and the error rate in positioning, as shown in Fig.4.3. This correlation provides valuable insights into the reliability of the estimated locations and paves the way for improving the robustness of localization tasks. By identifying and consequently filtering out the points associated with high uncertainty, we can considerably enhance the precision of subsequent location-based systems.

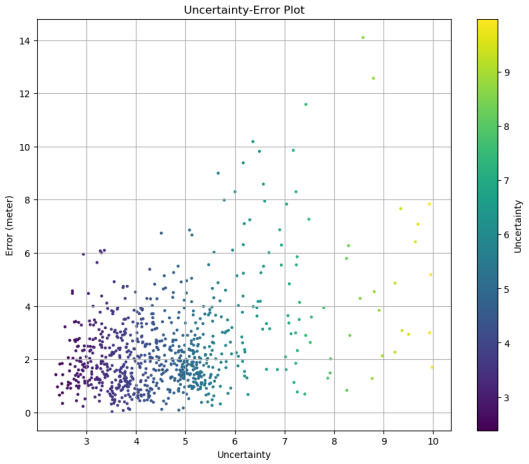


Figure 4.3: Estimated Uncertainty v.s. Localization Error

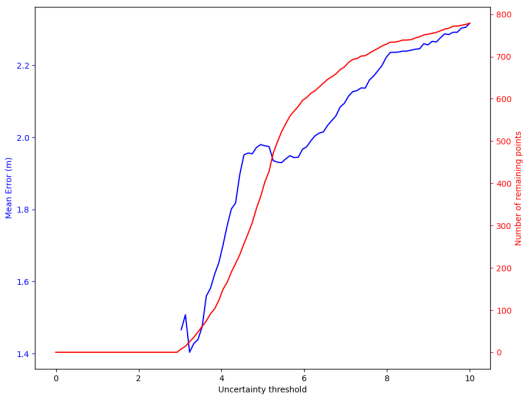


Figure 4.4: Filter threshold v.s. Mean error v.s. Points remaining

As evident from Fig. 4.4, even a minimal filtering of predictions based on uncertainty can lead to a substantial enhancement in the accuracy of the remaining predictions. This underscores the effectiveness of integrating uncertainty measures into the prediction process. In our experiments, an uncertainty threshold of 5.5 has been commonly utilized to achieve significant improvements in localization accuracy.

This refined filtering process enables more accurate and reliable positioning, minimizing the error propagation through successive systems in a given workflow.

4.2.3 Adaptive 'K' Selection via Uncertainty Estimation

In the context of WKNN-based WiFi localization, the selection of 'K' (i.e., the number of nearest neighbors considered in the algorithm) constitutes a critical aspect that directly relates to the quality and the density of the radio map database. When dealing with a sparsely populated radio map, it may be necessary to increase 'K' to gain a more accurate approximation of the correct general location at the expense of some degree of precision. Conversely, for densely populated radio maps, it might be beneficial to reduce 'K' to achieve higher localization precision. However, it's crucial to note that averaging over 'K' data points inevitably leads to the introduction of the inherent bias present within the radio map into the results.

Adjusting 'K' for each different environment, based on the associated radio map, can add an additional layer of complexity to the localization process. As addressed in the previous section, a methodology for estimating the uncertainty to gauge the localization error was introduced. Empirical evidence suggests that this uncertainty estimation can effectively facilitate the determination of the optimal 'K' value.

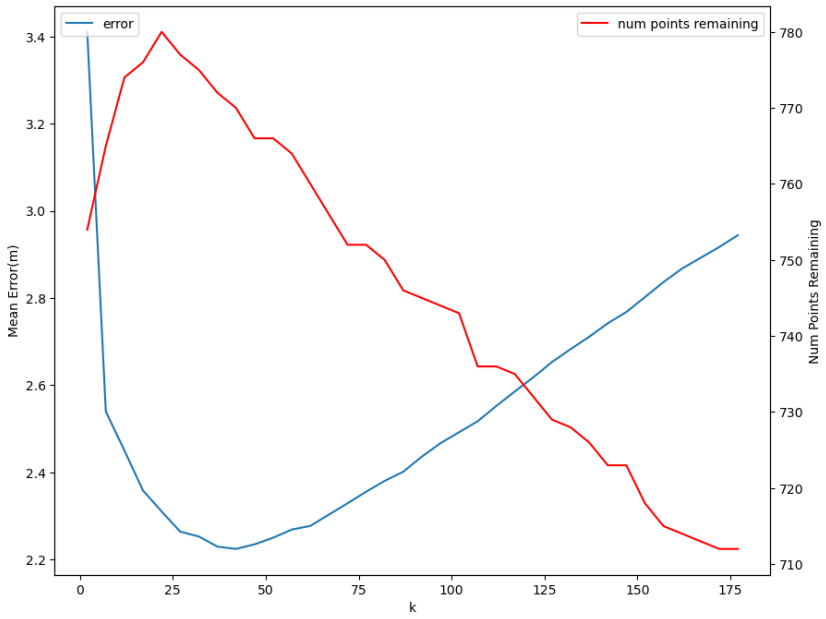


Figure 4.5: error v.s. points remaining v.s. K

Fig.4.5 demonstrates the relationship between the mean localization error and the quantity of data points with predefined uncertainty below 8 across a range of 'K' values. It can be observed that the selection of 'K' that drives a greater proportion of data points towards regions of low uncertainty tends to potentially ameliorate the overall mean accuracy of the localization process. This finding corroborates the proposition that optimal 'K' selection is pivotal in enhancing the localization precision and demonstrates the value of incorporating uncertainty estimation in the decision-making process.

We observe that simply selecting the 'K' corresponding to the highest remaining quantity of data points is sufficient to align 'K' closely with the minimum of the error curve. Another observation of interest pertains to the minimum error always present on the right side of the maximum number of remaining points. This phenomenon is likely attributable to the inherent summation and averaging characteristics of WKNN. From an engineering standpoint, it may be beneficial to incorporate an additional 10-20 points for enhanced performance. This strategy allows for the attainment of lower localization errors, further demonstrating the efficacy of employing an uncertainty-based methodology in the 'K' selection process in WKNN.

However, given the diverse nature of our dataset, we cannot assure the universal effectiveness of this approach across all environments. Consequently, in the ensuing results and comparisons, we continue to adopt a conventional 'K' value of 25.

4.3 Results and Discussion

4.3.1 Quantitative Analysis

We conducted a result filtering by utilizing estimated uncertainty, which, similar to Huawei's WKNN, involved the removal of the 90 most uncertain points. This new set of results was then evaluated with the performance of the aforementioned Huawei's WKNN for a more comprehensive evaluation.

Table 4.1: Comparison of Different Methodologies with 90 points removal

Methodology	rm Ept.	MAP	Mean	Median	Max	Percentile 95
Cosine			2.2738	2.0970	10.7795	4.5561
Manhattan			2.4597	2.0381	12.9379	5.4425
Euclid			2.6016	2.2675	13.8027	5.5129
Cosine	+		2.0872	1.8284	**9.1661**	4.4392
Manhattan	+		2.0535	1.8634	12.1802	4.1909
Euclid	+		2.1335	1.8263	11.8718	4.7863
Cosine	+	+	2.0805	1.8284	9.64356	4.3636
Manhattan	+	+	**2.0388**	1.8407	12.1802	**3.5938**
Euclid	+	+	2.1201	**1.8226**	11.8718	4.7635
Huawei	?	?	2.3625	2.0358	12.9366	5.2947

It can be observed from Table 4.1 that, with the exception of the Manhattan and Euclidean distance measurement methodologies without any Empty Removal or Radio Map Improvement, all other strategies outperform Huawei's WKNN method. Additionally, it is noteworthy that the cosine similarity measure exhibits a superior estimation of uncertainty. As a result, after uncertainty-based filtering, it consistently yields a smaller maximum error compared to other methodologies, thus effectively mitigating potential impacts on subsequent systems.

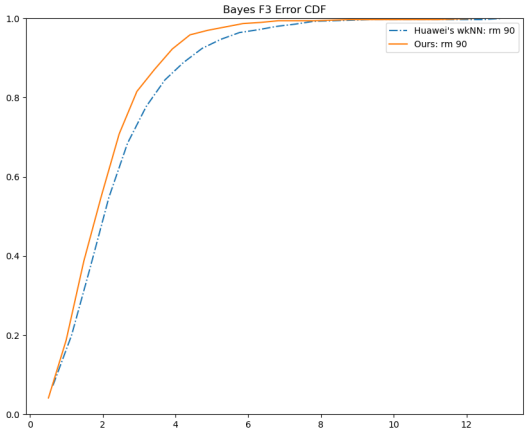


Figure 4.6: Error CDF

We also employ the common Error Cumulative Distribution Function (CDF) for comparative analysis as in Fig. 4.6. The CDF graph presented herein utilizes the Manhattan distance method, with both Empty Removal and Radio Map Improvement implemented. It is evident that our methodology holds an advantage over Huawei's approach across any given error range.

4.3.2 Visual Analysis

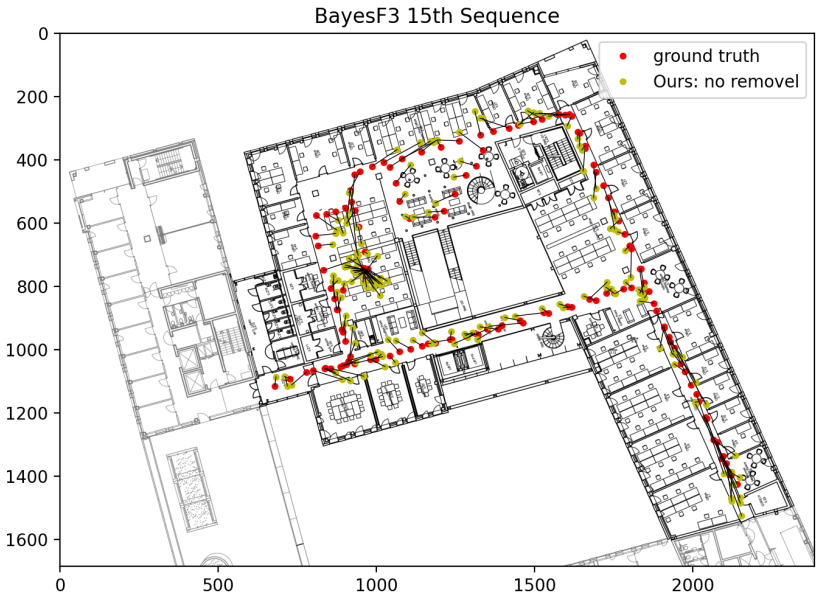


Figure 4.7: Visualization: ours

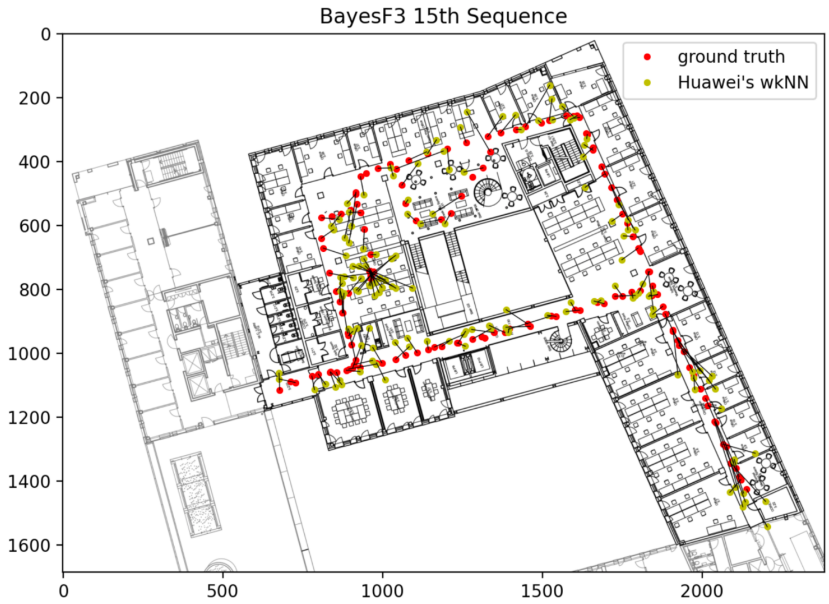


Figure 4.8: Visualization: Huawei

Figures 4.7 and 4.8 present visual representations of localization results for a specific sequence extracted from the Ground Truth (GT) data. These visuals intuitively display the actual locations and corresponding localization estimates, connected by lines for precise pairing identification. Comparatively, our results rarely venture into un-accessible regions, such as outside of corridors, demonstrating a higher degree of robustness than Huawei's results.

4.4 Summary

This chapter has introduced a method for estimating uncertainty in Weighted K-Nearest Neighbors (WKNN) localization, which has been validated to correlate with errors. By filtering out a fraction of points deemed most uncertain, a significant reduction in error has been achieved. Additionally, we've introduced a method for estimating the optimal value of 'K' using uncertainty. This approach intuitively selects the 'K' value that yields lower uncertainty across more points, serving as an effective approximation for the optimal 'K' value that minimizes error.

In the next chapter, we synthesize the concepts introduced in the preceding two chapters, culminating in the development of an autoencoder-based fingerprint embedding tailored for the WKNN approach, employing the Transformer architecture as the encoder.

Chapter 5

Fingerprint Auto-Encoder for WKNN Feature Extraction

5.1 Introduction

The previous chapters elaborated on a method of logarithmic scaling that capitalizes on the physical significance of signal propagation, providing a superior representation as evinced by the improved precision in Weighted K-Nearest Neighbors (WKNN) localization. However, there remains room for potential enhancement. The richness and volume of the data contained within the radio map signal an opportunity - one that invites the application of deep learning methodologies to derive an even more refined fingerprint representation. Harnessing the power of deep learning for feature extraction could yield a more nuanced, high-level understanding of the data, potentially resulting in an increase in localization precision.

The transformer[14], a celebrated mainstay in the realm of Natural Language Processing (NLP), brings unparalleled efficiency through its high degree of parallelizability and an escalating model capacity with additional layers. Moreover, the technique of word embedding prevalent in NLP offers an inherent methodology for representing individual Access Points (APs).

In this chapter, we introduce a method that exploits the strengths of the Transformer model to generate a more refined feature representation for fingerprints. The incorporation of word embedding techniques adapted from NLP to characterize each AP further enhances the sophistication of this representation. By bridging the principles of NLP and indoor localization, we strive to realize a more nuanced and effective fingerprint representation for improved localization performance.

5.2 Methodology

5.2.1 Transformer and APs' Word Embedding

In contrast to the conventional sequence-to-sequence transformation in the Transformer, our task is to transform a sequence of Access Points (APs), each with its MAC address and received signal strength (RSSI), into a dense feature vector, the Fingerprint Feature.

This transformation, rather than aiming for a complete translation between sequences, targets the extraction of a compact yet comprehensive representation from the input sequence. Similar to how large pre-trained language models like BERT utilize a special CLS token to encapsulate the semantic representation of an entire sentence [16], we also adopt a similar strategy by selecting a specific token from the output sequence to serve as the Fingerprint Feature representation.

In this context, each AP is considered analogous to a word in Natural Language Processing (NLP). A learnable word embedding is associated with each AP, effectively capturing its unique characteristics within the localization context. This not only provides a meaningful representation of the APs, but also facilitates learning from the patterns in their signal strengths, which are analogous to the contextual features of words in NLP.

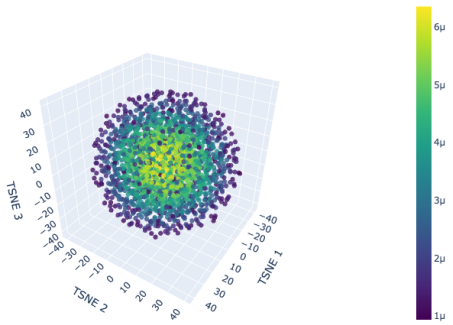


Figure 5.1: APs' 3D TSNE Embedding with Gaussian Kernel Density Estimation (KDE)

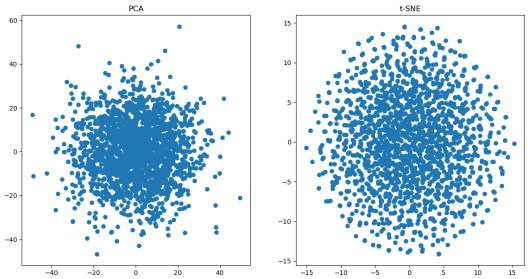


Figure 5.2: APs' 2D Embedding

The t-SNE (t-Distributed Stochastic Neighbor Embedding) [17] and PCA (Principal Component Analysis) [18] plots depicted in Figures 5.2 and 5.1 provide insightful visualizations of the learned embeddings. These plots resemble characteristics commonly seen in natural language word embeddings: one region appears more densely populated, while others are relatively sparse.

This distribution of embeddings in the vector space can profoundly influence the results of the self-attention mechanism in the Transformer. For instance, if a subset of APs often co-occur or are located close to each other, their embeddings might be closer in the vector space. This proximity in the embedding space can reflect the underlying patterns of the data, such as the physical locations of APs or their co-occurrence in fingerprints.

In natural language processing applications, Transformers typically require a positional encoding to maintain the order of the sequence, reflecting the innate sequential nature of language. However, our task inherently lacks such an ordered structure. Specifically, the access points (APs) in a fingerprint don't maintain a consistent order that carries meaningful information, as the sequence order does not mirror their physical arrangement in the environment.

Consequently, we have eliminated the need for positional encoding in our adaptation of the Transformer. This approach allows the model to focus on the relationships between the APs as represented in the embeddings, rather than the arbitrary order in which they might appear in a sequence.

5.2.2 Model Structure

The architecture of our model emanates from modifications to the structure delineated in Chapter 3. Given the preceding exposition on this design, we will succinctly outline its evolution herein.

To model RSSI and AP within a single token, we initially explored the approach of creating another vocabulary set for RSSI values and concatenating two subtokens (one for AP and one for RSSI). However, we found this approach prone to overfitting due to the exponential increase in the size of the vocabulary. Ultimately, we adopted a methodology that directly incorporates RSSI values into the computation of attention scores, providing a more effective and less complex approach to the problem.

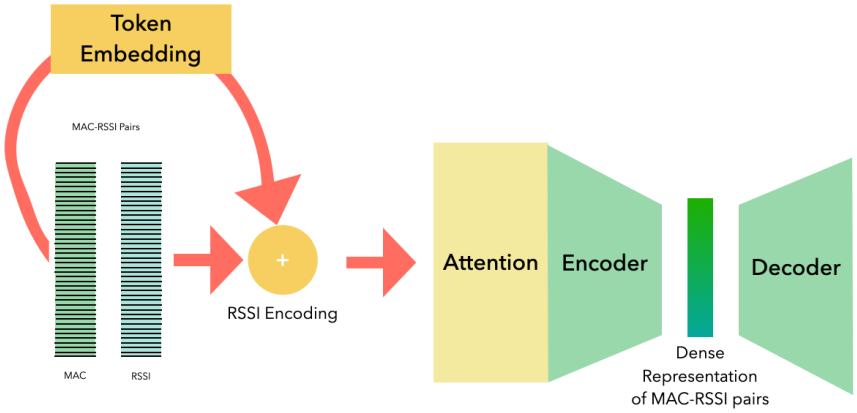


Figure 5.3: Model Architecture

The first output token from the self-attention outputs is harnessed as the input to the decoding phase. This decoder is architecturally composed of a series of fully connected neural networks, with the intended target of the decoding process being the RSSI vector.

This method finds parallels in the schema we adopted in the WKNN segment. There, we deliberately attenuated the prominence of signals exhibiting lower RSSI values during their transformation. This deliberate modulation not only integrated RSSI nuances into our model but also ensured that each AP's embedding underwent logarithmic scaling in congruence with its specific RSSI magnitude.

The Transformer's attention mechanism can be described using the following equations, as commonly represented in the original Transformer paper by Vaswani et al. [14]:

5.2.3 Auto-Encoder Training

In our aim to extract a dense FP (Fingerprint) embedding, we require a task to train our model. We have chosen an Auto-Encoder for this purpose, as shown in Fig. 5.3. The task of the model is to reconstruct the input FP, which is also log-scaled as introduced in the WKNN section, forming a one-dimensional vector with a length equivalent to the total number of APs. The utilization of Auto-Encoder aids in the development of a compact and meaningful representation of FP, enabling the model to capture the most salient features of the input while minimizing scaled information loss.

The configuration for our model consists of an input token length of 128. Each input sequence will be padded to a length of 200 tokens. After eight layers of self-attention, the transformer outputs a one-dimensional vector of length 128, encapsulating all FP

information. This dense vector is then decoded by a series of linear layers.

For our loss function, we commonly employ Mean Squared Error (MSE) or Huber Loss, in view of their respective strengths in managing outliers and ensuring a smooth error surface than L1 Loss. The AdamW [19] optimizer is used for training the model, with hyperparameters set to a learning rate of 1e-4, beta1 = 0.9, and beta2 = 0.999. These configurations have been found to yield reliable results across a variety of environments and hence were employed in this context.

Additionally, to enhance the robustness of our model and diminish the dependence on specific Access Points (APs), we have incorporated noise into our training regime, thus validation loss is always lower than training loss as shown in Fig. 5.4. This is done by implementing a dropout mechanism, whereby 20% of APs in each input sample are consistently deactivated during the training phase. Despite this, the model is still required to reconstruct the entire set of APs. This dropout strategy encourages the model to learn a more generalized and stable representation of the FP information, which in turn improves its capacity to handle real-world uncertainties and potential data discrepancies.

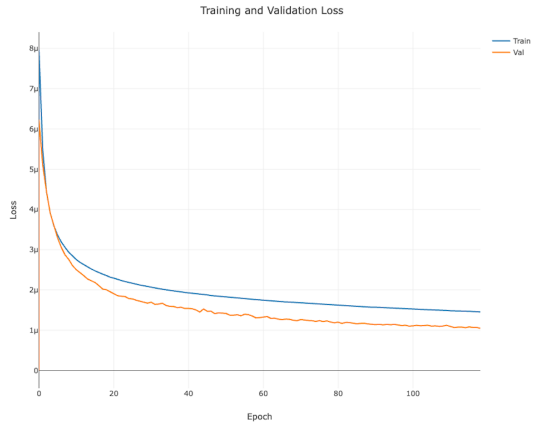


Figure 5.4: Training Curve

5.2.4 Use Coded Embeddings as Features For WKNN

This unsupervised training of this model results in a dense embedding that can be utilized for numerous tasks, with localization being just one of the feasible applications. We largely adhere to the principles of the Weighted K-Nearest Neighbors (WKNN) methodology as shown in Alg. 2, but the crucial distinction is that the feature engineering process, traditionally performed manually, has now been automated. This automation has the potential to discover more intricate patterns within the data, leading to more

precise and comprehensive feature representations, thereby enhancing the performance of localization tasks.

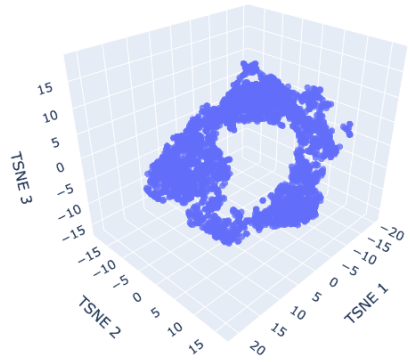


Figure 5.5: Bayes F3 FP Embeddings

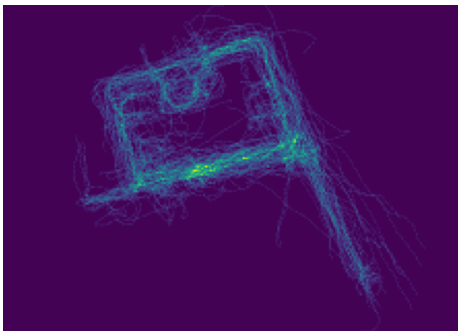


Figure 5.6: Bayes F3's Radio Map

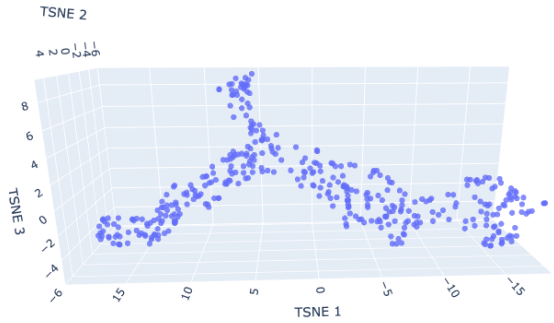


Figure 5.7: St. James FP Embeddings

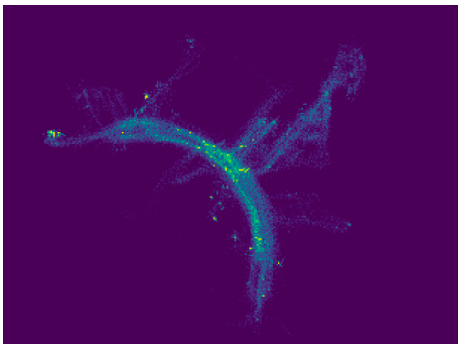


Figure 5.8: St. James's Radio Map

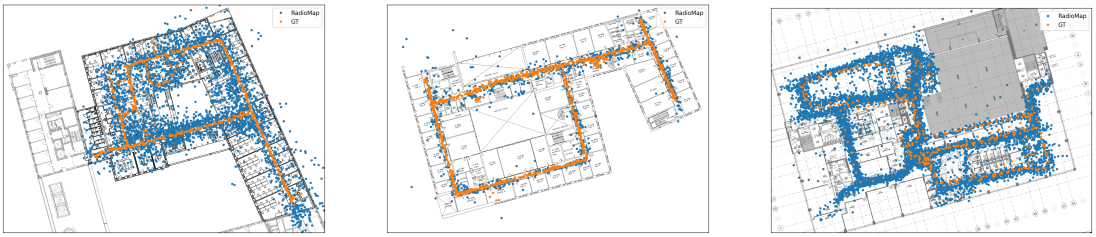
The above figures depicts the t-SNE 3D visualization of the Fingerprint (FP) embeddings and their radio map within 2 buildings. Notably, these embeddings remain continuous in high-dimensional space and bear resemblance to the physical layout of the building itself, reflecting the structure of the radio map, including features like the shape of the corridors. However, it is essential to underline that the model was not fed with any explicit positional information during training. The fact that the model could infer spatial characteristics on its own is a fascinating discovery. This unsupervised learning of spatial features demonstrates the transformative potential of deep learning models in mapping unstructured environments, such as indoor settings, even without location-specific input. Future work can explore the depth and boundaries of this spontaneous spatial inference and potentially use it to enhance positioning accuracy further.

In addition to crafting general fingerprint representations, a unique aspect of our approach is the extraction of individual Access Point (AP) embeddings as shown in Fig. 5.2. This feature is not commonly found in most other methodologies in the field. By obtaining AP-specific embeddings, we are enabled to reverse-engineer the properties of individual APs, such as their coverage range or location within the building structure. This valuable information could be utilized for optimizing the deployment of APs, improving overall network performance, and facilitating more effective management of the wireless environment.

Algorithm 2 Fingerprint Embedding for WKNN Positioning

procedure WKNN($FP_t, FP_{\text{ref}}, k, a$)
    $Feature_t \leftarrow$ Pre-trained Model(FP_t)
    $Feature_{\text{ref}} \leftarrow$ Pre-trained Model(FP_{ref})
    Initialize an empty list Δ
    for each FP_r in FP_{ref} **do**
        Compute $\Delta_{t,r}$ using $criterion(Feature_t, Feature_{ref})$ (either Cosine or Manhattan)
        Add $\Delta_{t,r}$ to Δ
    end for
    Sort Δ in ascending (or descending) order and select the top nearest k entries to get $\Delta_{\text{top\_k}}$
    Identify the positions $Pos_{\text{top\_k}}$ corresponding to $\Delta_{\text{top\_k}}$ in $Feature_{\text{reference}}$
    Compute $Pos_t = \frac{[Pos_1, Pos_2, \ldots Pos_k] \cdot [\Delta_1, \Delta_2, \ldots \Delta_k]}{k \cdot \sum_{i=1}^{k} \Delta_i}$
    return Pos_t
end procedure

5.3 Results

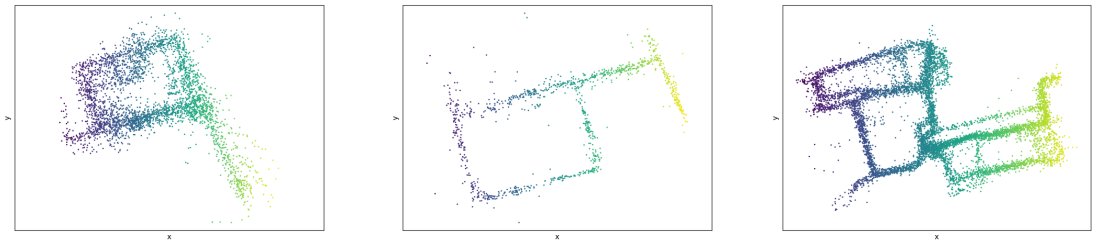


(a) Bayes F3 (b) Forum F1 (c) Main Library

Figure 5.9: Radio Map & GT

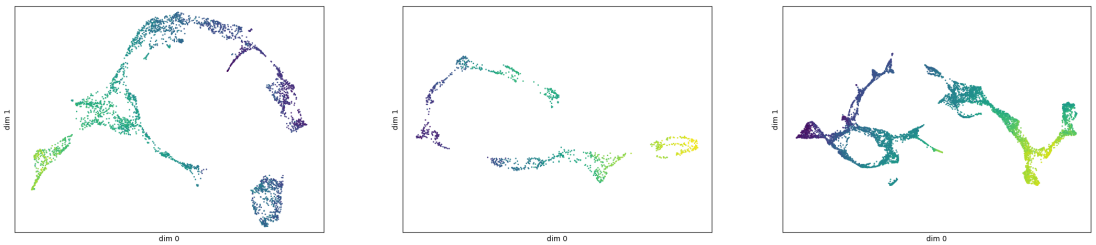
(a) Bayes F3 (b) Forum F1 (c) Main Library

Figure 5.10: Positioning Output



(a) Bayes F3 (b) Forum F1 (c) Main Library

Figure 5.11: Colorized Radiomap



(a) Bayes F3 (b) Forum F1 (c) Main Library

Figure 5.12: Colorized Embedding Visualization

Methods	Bayes F3			Forum F1			Main Libiray F1		
	Mean	Median	Max	Mean	Median	Max	Mean	Median	Max
Naive-WKNN	3.0009	2.45717	11.1487	3.6371	3.1127	15.5875	2.6733	2.1464	14.5017
Regression	2.2762	1.91042	14.4922	3.5603	2.8370	21.1554	2.65971	2.340926	12.0070
Neural WKNN	2.1157	1.6983	11.2422	2.5872	2.12141	15.2917	2.42252	2.0340	13.9391

Table 5.1: Comparison results in positioning error in meters with the other 2 methods

To rigorously assess the locational performance of our model, we conducted comprehensive experiments across three distinctive domains: Bayes F1, Forum F1, and Main Library F1. These environments, each with its inherent challenges and characteristics, offer a holistic canvas to gauge the efficacy of various positioning methods.

The results, as illustrated in the table 5.1, underscore the preeminence of the AE-based Neural WKNN approach in terms of locational precision. This approach consistently outperforms both the Regression and the Naive WKNN methods across all testing grounds.

Bayes F1: The Neural WKNN manifests a notably reduced mean error of 2.1157 meters, compared to the Naive WKNN's 3.0009 meters and Regression's 2.2762 meters. Furthermore, the maximum error and median error are appreciably lower, emphasizing the robustness of the Neural WKNN in this environment.

Forum F1: While all three methods have their merits, the Neural WKNN continues its dominant performance with a mean error of 2.5872 meters. Notably, it maintains a relatively compact range between its median and maximum error, signifying consistent accuracy.

Main Library F1: Neural WKNN showcases a commendable balance between mean and median errors, underscoring its stable performance. With a mean error of 2.42252 meters, it outperforms the Naive WKNN and closely rivals the Regression model.

The findings indicate that the Neural WKNN outperforms both the conventional WKNN and Regression techniques. The inherent noise in WiFi signals is effectively mitigated by the neural network, while the positional noise is neutralized by the amalgamation of multiple WKNN candidates.

A series of figures, Fig5.5 to Fig5.8, provide a comprehensive visualization of the datasets employed, the model's performance, and intriguing aspects of the radio map and its transformations.

Fig5.5: This figure offers a juxtaposition of the training and testing datasets. A cursory glance reveals the stark contrast between the Radio Map, which is rife with perturbations, and the Ground Truth (GT) which is remarkably regular. This discrep-

ancy illuminates one of the key reasons for the superior performance of WKNN over regression; the underlying complexity and noise inherent in the Radio Map necessitate a method robust enough to navigate and interpret its intricacies.

Fig5.6: The results of Neural WKNN's positioning efforts are showcased here. There is a conspicuous regularity in the locational predictions. This regularity, induced by the incorporation of K-neighbors, signifies the neutralization of a significant portion of the noise present in the Radio Map, highlighting the efficacy of the WKNN approach.

Fig5.7: Serving as a precursor to Fig5.8, this figure portrays the radio map with colors denoting the x and y coordinates. The primary aim of this visualization is to set a baseline for the ensuing comparison in Fig5.8.

Fig5.8: This figure employs the UMAP dimensionality reduction tool to present a 2D visualization of the embeddings from the autoencoder (AE) used in our model. One of the salient observations from this visualization is the preservation of neighborhood integrity. Points that are proximate in the embedding space also retain their closeness in the Radio Map's coordinates. Even more fascinating is the model's inherent ability to discern and replicate the topological nuances of the building structures. The embedded shapes, which exhibit a marked resemblance to the actual structures, attest to the model's capacity to autonomously comprehend the architectural topography.

In addition to the aforementioned comparisons, we extended our evaluation to encompass industrial-grade algorithms, specifically one from Huawei. Particularly, in the context of the Bayes F3 dataset.Through the representation provided by the Cumulative Distribution Function (CDF), it is evident that our model outperforms the industrial-grade algorithms.

The visualization of the estimated uncertainty and its subsequent filtering process are depicted in Fig. 5.14 and 5.15. The supplementary uncertainty estimation empowers our approach further, enabling the filtration of ambiguous points, thus offering robust support for subsequent data fusion.

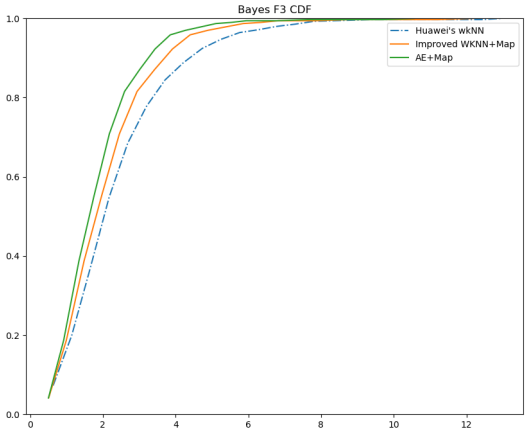


Figure 5.13: CDF comparison

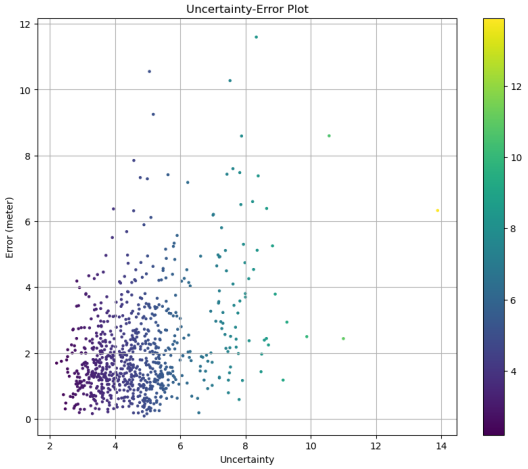


Figure 5.14: Estimated Uncertainty

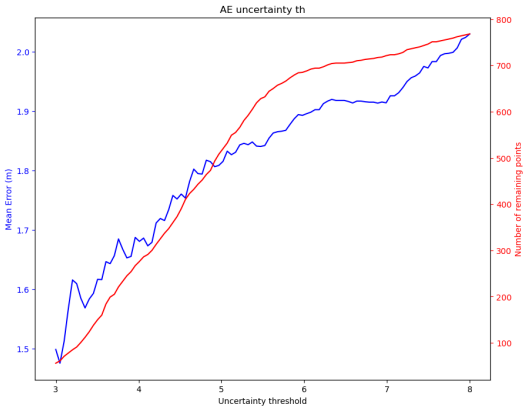


Figure 5.15: Uncertainty Filtering

In essence, these visual analyses and quantitative results not only validate the superiority of Neural WKNN but also spotlight its adeptness at unravelling and leveraging

the spatial semantics embedded in the Radio Map for precise localization.

5.4 Summary

This chapter presents a technique harnessing the considerable inferential capabilities of the Transformer autoencoder model. This approach yields superior localization accuracy compared to regression methods, as it adeptly captures both the features of Access Points (AP) and the distinctive characteristics of the Radio Map within a specific region. By harnessing its powerful inferential capabilities, we managed to seamlessly integrate the features of Access Points (AP) and the unique attributes of Radio Maps within specific regions. This integration yielded a feature representation that was both robust and general, enhancing our model's performance in localization tasks significantly.

A unique aspect of our model is its engine's reliance on the Weighted K-Nearest Neighbors (WKNN) algorithm. This approach not only retains the established advantages of WKNN but also supplements them with the advanced feature extraction prowess of the Transformer AutoEncoder. Notably, the Uncertainty estimation method, introduced in Chapter 4, remains valid and is further fortified by this synergistic combination.

Despite the model's strengths, it's worth considering the broader implications and limitations of such an approach. A powerful feature representation of WiFi characteristics within a specified area undoubtedly provides advantages. However, from an industrial perspective, the question becomes: Is there a universal model that could span across all buildings, sacrificing a slight decrease in precision for increased efficiency and eliminating the constant need for training?

Future endeavors should probe into these pertinent questions. The balance between model complexity, its generalizability across diverse environments, precision, and computational efficiency is a fertile area of investigation. Another promising avenue could be the adoption of transfer learning techniques, allowing models to leverage learnings from one building or environment to another, optimizing both efficiency and accuracy.

Chapter 6

Conclusion

In the ever-evolving domain of WiFi RSSI positioning, the need for robust, efficient, and precise localization techniques is paramount. This dissertation has ventured deep into this realm, exploring and establishing novel methodologies that leverage the strengths of deep learning, the foundational principles of WKNN, and the intricacies of uncertainty estimation.

Our dual-pronged approach, rooted in Deep Evidential Learning theory and WKNN candidate variance, has not only delivered precision surpassing industry benchmarks but also introduced the concept of Neural WKNN. Utilizing Transformer-based autoencoders, we've carved out robust embeddings that crucially augment localization accuracy.

However, central to our narrative is the nuanced understanding and estimation of uncertainty. While accuracy is undeniably a cornerstone, it's the intelligent treatment of uncertainty that sets our methodology apart. This became particularly evident through empirical tests conducted across the University of Edinburgh.

Looking ahead, there's immense potential in refining the metrics of uncertainty even further. Future endeavors should aim at deriving a more precise measure of uncertainty and imbuing it with tangible physical significance. Such advancements would facilitate smoother data integration steps in WiFi-IMU sensor fusion.

Bibliography

[1] Navigine. *WiFi for Indoor Positioning and Navigation*. 2023. URL: https://navigine.com/blog/wifi-for-indoor-positioning-and-navigation/.

[2] AZoSensors. 'New WiFi Technique Helps Indoor Robot Navigation'. In: *AZoSensors* (2021). URL: https://www.azosensors.com/news.aspx?newsID=15051.

[3] S. Woo, S. Jeong, E. Mok et al. 'Application of WiFi-based indoor positioning system for labor tracking at construction sites: A case study in Guangzhou MTR'. In: *Automation in Construction* 20.1 (2011). Global convergence in construction, pp. 3–13. ISSN: 0926-5805. DOI: https://doi.org/10.1016/j.autcon.2010.07.009.

[4] R. Focus. *How WiFi and Indoor Location Marketing are Changing the Game in Retail*. 2020. URL: https://retail-focus.co.uk/how-wifi-and-indoor-location-marketing-are-changing-the-game-in-retail/.

[5] X. Feng, K. Nguyen and Z. Luo. 'A survey of deep learning approaches for WiFi-based indoor positioning'. English. In: *Journal of Information and Telecommunication* (Sept. 2021). ISSN: 2475-1839. DOI: 10.1080/24751839.2021.1975425.

[6] K. Kostas, R. Y. Kostas, F. Zampella and F. Alsehly. *WiFi Based Distance Estimation Using Supervised Machine Learning*. 2022. arXiv: 2208.07190 [cs.LG].

[7] N. Singh, S. Choe and R. Punmiya. 'Machine Learning Based Indoor Localization Using Wi-Fi RSSI Fingerprints: An Overview'. In: *IEEE Access* 9 (2021), pp. 127150–127174. DOI: 10.1109/ACCESS.2021.3111083.

[8] M. Abbas, M. Elhamshary, H. Rizk, M. Torki and M. Youssef. 'WiDeep: WiFi-based Accurate and Robust Indoor Localization System using Deep Learning'. In: *2019 IEEE International Conference on Pervasive Computing and Communications (PerCom*. 2019, pp. 1–10. DOI: 10.1109/PERCOM.2019.8767421.

[9] A. Belmonte-Hernández, G. Hernández-Peñaloza, D. Martín Gutiérrez and F. Álvarez. 'Recurrent Model for Wireless Indoor Tracking and Positioning Recovering Using Generative Networks'. In: *IEEE Sensors Journal* 20.6 (2020), pp. 3356–3365. DOI: 10.1109/JSEN.2019.2958201.

[10] M. T. Hoang, B. Yuen, X. Dong, T. Lu, R. Westendorp and K. Reddy. 'Recurrent Neural Networks for Accurate RSSI Indoor Localization'. In: *IEEE Internet of Things Journal* 6.6 (2019), pp. 10639–10651. DOI: 10.1109/JIOT.2019.2940368.

[11] M. T. Hoang, B. Yuen, X. Dong, T. Lu, R. Westendorp and K. Reddy. 'Recurrent Neural Networks for Accurate RSSI Indoor Localization'. In: *IEEE Internet of Things Journal* 6.6 (2019), pp. 10639–10651. DOI: 10.1109/JIOT.2019.2940368.

[12] J. Devlin, M.-W. Chang, K. Lee and K. Toutanova. *BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding*. 2019. arXiv: 1810.04805 [cs.CL].

[13] A. Amini, W. Schwarting, A. Soleimany and D. Rus. *Deep Evidential Regression*. 2020. arXiv: 1910.02600 [cs.LG].

[14] A. Vaswani, N. Shazeer, N. Parmar et al. *Attention Is All You Need*. 2017. arXiv: 1706.03762 [cs.CL].

[15] A. Foliadis, M. H. C. García, R. A. Stirling-Gallacher and R. S. Thomä. 'Reliable Deep Learning based Localization with CSI Fingerprints and Multiple Base Stations'. In: *CoRR* abs/2111.11839 (2021). arXiv: 2111.11839. URL: https://arxiv.org/abs/2111.11839.

[16] J. Devlin, M.-W. Chang, K. Lee and K. Toutanova. *BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding*. 2019. arXiv: 1810.04805 [cs.CL].

[17] L. van der Maaten and G. Hinton. 'Viualizing data using t-SNE'. In: *Journal of Machine Learning Research* 9 (Nov. 2008), pp. 2579–2605.

[18] A. Maćkiewicz and W. Ratajczak. 'Principal components analysis (PCA)'. In: *Computers Geosciences* 19.3 (1993), pp. 303–342. ISSN: 0098-3004. DOI: https://doi.org/10.1016/0098-3004(93)90090-R. URL: https://www.sciencedirect.com/science/article/pii/009830049390090R.

[19] I. Loshchilov and F. Hutter. *Decoupled Weight Decay Regularization*. 2019. arXiv: 1711.05101 [cs.LG].