Multiview Variational Deep Learning with Application to Scalable Indoor Localization

Minseuk Kim  
School of Electrical Engineering  
KAIST, Daejeon, Republic of Korea  
kms4105@kaist.ac.kr

Changjun Kim  
School of Electrical Engineering  
KAIST, Daejeon, Republic of Korea  
changjun0605@kaist.ac.kr

Dongsoo Han  
School of Computing  
KAIST, Daejeon, Republic of Korea  
dshan@kaist.ac.kr

June-Koo Kevin Rhee  
School of Electrical Engineering  
ITRC of Quantum Computing for AI  
KAIST, Daejeon, Republic of Korea  
rhee.jk@kaist.edu

Abstract

Radio channel state information (CSI) measured with many receivers is a good resource for localizing a transmit device with machine learning with a discriminative model. However, CSI localization is nontrivial when the radio map is complicated, such as in building corridors. This paper introduces a view-selective deep learning (VSDL) system for indoor localization using CSI of WiFi. The multiview training with CSI obtained from multiple groups of access points (APs) generates latent features on supervised variational deep network. This information is then applied to an additional network for dominant view classification to enhance the regression accuracy of localization. As non-informative latent features from multiple views are rejected, we can achieve a localization accuracy of 0.77 m, which outperforms by 30% the best known accuracy in practical applications in a real building environment. To the best of our knowledge, this is the first approach to apply variational inference and to construct a scalable system for radio localization. Furthermore, our work investigates a methodology for supervised learning with multiview data where informative and non-informative views coexist.

1 Introduction

Machine learning applications with a multiview embodiment rendered with multiple sources can exploit feature correlations among various views to attain the best model for inference. Multiview data can be efficiently generated by variational inference \[1\] from a single view, as reported in \[2\]. Variational inference is also adopted in broad discriminative models such as clustering \[3\], classification \[4\], and regression \[5\] to utilize the probabilistic latent feature space. From the perspective of a deep network, variational deep learning (DL) \[6\] can jointly optimize objectives of a Gaussian process marginal likelihood to train the deep network. Reparameterization of variational inference \[7\] derives mean-field latent feature vectors to represent the posterior of an input. Compared to stand-alone deep networks, support vector machines (SVMs), and other recent systems, the variational DL can substantially improve classification and regression performance \[7\]. In this paper, we apply the variational DL to localization of transmit devices based on radio signal data in a practical WiFi environment.

Similarly to an example that generates a single image view from multiple image views \[8\], radio signals retrieved in multiple groups of receivers at distributed locations can form multiview information...
for machine learning to locate the transmit device. By the nature of variational DL that the elements of
the extracted latent vector have no correlations, we can map between a multivariate standard normal
distribution and each view. As the variational DL is forced to ensure its latent vector approximated to
the standard normal distribution with a reduced number of independent Gaussian processes, we can
mitigate the signal uncertainties. Compared to the global positioning system (GPS), which leads to 5
m to 10 m outdoor localization accuracy, indoor localization requires more accurate positioning of
the transmit devices. Recently, a channel state information (CSI) of WiFi has emerged as the strong
candidate for indoor localization rather than a scalar-valued received signal strength indicator (RSSI)
[9]. The subcarrier CSIs of WiFi orthogonal frequency division multiplexing (OFDM) channel form
complex vector information, providing much rich information on radio localization with an excellent
localization accuracy.

1.1 Related work
At a server side of a localization system, CSI data is collected from multiple receivers at the same
time to find the transmit device location. Transmit device localization is achieved by geometric
analysis of such CSI data to determine the time of flight and angle of arrival information of the radio
packet [10, 11]. However, in a practical indoor environment, the noise and signal fading problems
become critical against finding true transmit device locations with such analytical methods.

In order to cope with noise and signal fading, many machine learning methods have been developed
to successfully find the transmit device location from the complex CSI data by considering it as a
single view. References [12, 13, 14] utilized restricted Boltzmann machine (RBM) based approaches
to reconstruct the CSI data for better likelihood determination for localization. Convolutional neural
network (CNN) based approaches were proposed in [15, 16, 17]. The consecutive data packets were
concatenated to a single batch as a 2-D CSI image. However, in the radio localization, it is preferable
that the system be capable of packet-by-packet processing, rather than batch processing. References
[18] and [19] introduced SVM based classification and regression, respectively. Reference [20]
adopted transfer learning to reconstruct CSI data and applied the enhanced k-nearest neighbor (KNN)
approach for localization. To enhance the accuracy of spot location classification, [21] introduced
autoencoder. In [22], principal component analysis (PCA), one of the preprocessing methods, was
applied to reduce the multi dimension CSI before passing through a deep neural network (DNN).
Also, combining RSSI and CSI, [23] proposed multi-layer perceptron (MLP) and 1-D CNN. From
the perspective of device-free indoor localization, [24] carried out a canonical correlation analysis
(CCA) to classify the location of a device with a human, where the device is neither a transmitter nor
a receiver. The localization accuracies achieved by all the papers listed above did not reach below a
meter in practical application environments, except for the case where the training and test locations
are the same. However, our result in this paper has achieved a localization accuracy of sub-meter by
adopting multiview architecture with variational DL.

1.2 Our contributions
An advanced scalable localization with a keen accuracy is pursued in this work, to extend the area of
radio localization on a complicated floor plan for an in-building application. To construct a scalable
learning system for localization in a real building environment with corridors, we propose a supervised
learning system named view-selective deep learning (VSDL) with CSI data consisting of multiple
views. The VSDL obtains much improved regression performance due to latent feature generation by
the use of the variational inference and non-informative latent feature rejection among the multiple
views. The proposed VSDL achieved a localization accuracy of 0.77 m, which outperforms by more
than 30% the best known accuracy of other works applied in practical building experiment. To the
best of our knowledge, this is the first approach to apply variational inference for CSI-based WiFi
localization and to construct a scalable system for localization in a wide and complex environment.
Furthermore, application of our system can be extended to general supervised learning with multiview
data where informative and non-informative views coexist.

2 CSI Preliminaries and data collection
In a WiFi network, device localization can be achieved by analyzing the CSIs of a radio packet
arriving at multiple receiver antennas of an access point (AP), complying with the IEEE 802.11a/g/n/ac
We apply a variational DL for regression to be trained with input as a pair of $x$ where a collection of APs in each sub-area forms a view of the CSI sample. We apply deep learning where a subset of subcarriers $\{q_i\}$ variational inference, the CSI is mapped to an unit phasor complex that measures phase difference of vector of 30 selected subcarriers for an antenna receiving a WiFi packet [25]. The phase difference of $q_i$ Our aim is to approximate the estimated posterior $q(z)\|\|p(z)$, where the true $y$ output from the latent vector $z$ indicate mean and standard deviation, respectively. The latent vector is used to estimate the regression $z$ of $x$ consisting of $y$ samples, defined as $x = [x_{m,i}]$, with $m \in \{1, \ldots, M-1\}$ and $i \in \{1, \ldots, I\}$.

In order to construct the localization system scalable in a complex area, we should consider exclusion of non-informative CSI views. One can deploy multiple APs over the area consisting of $K$ sub-areas, where a collection of APs in each sub-area forms a view of the CSI sample. We apply deep learning to classify the dominant views from multiple sub-areas.

3 Variational Deep Learning

We apply a variational DL for regression to be trained with input as a pair of $x$ (i.e., CSI in our case) and a true label $y$ (i.e., Cartesian coordinate in our case). Let us assume a latent feature vector $z$ consisting of $z_j = f_j(x)$, $j \in \{1, \ldots, J\}$ of independent Gaussian processes (GPs) with probability of $z_j \sim GP(\mu_j, \sigma^2_j)$, where the mean-field feature vectors $\mu = [\mu_1, \ldots, \mu_J]$ and $\sigma = [\sigma_1, \ldots, \sigma_J]$ indicate mean and standard deviation, respectively. The latent vector is used to estimate the regression output $\hat{y}$ where the true $y$ is supposed to be represented as $y(x) = N(y(x); z, \sigma^2 \circ I)$. In order to make the learning variables differentiable for gradient descent based back-propagation, it requires reparameterization. The weights and biases of a neural network (NN) to obtain the latent vector $z$ from the input $x$ is updated by sampling of noise vector $\epsilon$:

$$z = \mu + \sigma \circ \epsilon, \quad \epsilon \sim N(0, 1).$$

Then with the variational posterior over the estimated distribution $q(z)$, Jensen’s inequality can be applied to give evidence lowerbound (ELBO) of the marginal log-likelihood of regression [6]:

$$\log p(y) \geq E_q(z|y)[\log p(y|z)] - KL[q(z|y)||p(z)] \triangleq \mathcal{L}(q) = \text{ELBO}. \quad (4)$$

In [4], the first term at the right hand side is cross entropy between the label $y$ and the regression output from the latent vector $z$, which is equivalent to the regression loss. The second term is the Kullback–Leibler (KL) divergence between the estimated posterior $q(z|y)$ and the distribution $p(z)$. Our aim is to approximate the estimated posterior $q(z|y)$ to be close to the posterior $p(z|y)$; in
other words, to minimize $\text{KL}[q(z|y)||p(z|y)]$. Again the log-likelihood can be represented as the following equation:

$$\log p(y) = \int \log (p(y)) q(z|y) dz$$

$$= \int \log \frac{p(y, z)}{q(z|y)} dz + \int \log \frac{q(z|y)}{p(z|y)} q(z|y) dz$$

$$= \text{ELBO} + \text{KL}[q(z|y)||p(z|y)].$$

Since the log-likelihood $\log p(y)$ is bounded, in order to minimize the KL divergence $\text{KL}[q(z|y)||p(z|y)]$, we have to maximize the term ELBO. The KL divergence term in (4) with the normal distribution $q(z|y)$ and the standard normal distribution $p(z)$ can be simplified to

$$\text{KL}[q(z|y)||p(z)] = \frac{1}{2} \sum_{j=1}^{J} (\mu_j^2 + \sigma_j^2 - \ln(\sigma_j^2) - 1).$$

By updating the NN parameters to jointly reduce the cross entropy and the KL divergence $\text{KL}[q(z|y)||p(z)]$, we can both approximate the posterior and estimate the desired regression output.

4 Localization: View-Selective Deep Learning

We introduce novel learning system, or the VSDL, with which the relative degree of importance of each view among the multiple views is used to improve the regression accuracy. A type of selective sampling, referred to as co-testing, was introduced [28] to efficiently extract features from multiview data, where its basic idea is to inject divided views to multiple independent learning networks. However, this model had a strict assumption that each data sample might have a strong correlation among the views. On the other hand, we focus on a situation where views are not correlated to each other nor informative, but there is a given information whether or not a view is informative. In our case, we require localizing a target in a two-corridor environment, as in Figure 1.

Figure 1: Indoor localization on a two-corridor environment. The radio signals from AP3 to A and AP1 to B are not on LoS. In terms of data views of APs, these NLoS signals make the corresponding views non-informative.

As you can see the Figure 1, the WiFi radio signal can propagate not only through a line of sight (LoS) but also non-line of sight (NLoS) paths. Subsequently the received signal consists of multi-path fading as well as the signal noise. The CSI data view at APs only from NLoS paths are non-informative, since the multi-paths are very unpredictable. In this kind of case, it must be inefficient if all views are included in training. However, we can judge whether or not a view is informative so as to utilize such given information for supervised training. In our VSDL model, the learning parameters are much effectively updated by excluding non-informative data views. Looking into the big picture, VSDL is designed as a two-stage learning network for regression, consisting of a view-oriented variational deep network and a view-classified regression network.

4.1 View-Oriented Variational Deep Network

The VSDL model is first trained to extract latent feature vectors from the multiview data in a view-oriented way, as shown in the left grey box in Figure 2. We first reconstruct the multiview input data
To apply variational inference, each latent sampling NN encodes \( x_k \) to latent feature \( z_k \) by the mapping, \( z_k = h_k(x_k) \), with optimized mean-field vectors \( \mu_k = [\mu_{1,k}, \ldots, \mu_{J,k}] \) and \( \sigma_k = [\sigma_{1,k}, \ldots, \sigma_{J,k}] \) through \( L \) hidden layers, \( h_k^{(1)}, \ldots, h_k^{(L)} \). The weight and bias \( W_k^{(l)} \) and \( b_k^{(l)} \) of the layer \( l \in \{1, \ldots, L\} \) are used to evaluate the feature output \( \phi_k^{(l)} (\phi_k^{(0)} = x_k) \) for the next layer input. The output layer \( h_k^{(L)} \) generates the mean-field vectors:

\[
[\mu_k, \sigma_k] = W_k^{(L)} \phi_k^{(L-1)} + b_k^{(L)},
\]

followed by reparameterization for latent vector \( z_k \) as

\[
z_k = \mu_k + \sigma_k \odot \epsilon_k, \quad \epsilon_k \sim \mathcal{N}(0,1).
\]

Here, we try to minimize KL divergence of \( z_k \) according to (6) to approximate the posterior \( q(z_k | y) \) to the distribution \( p(z_k) \).

Along with the process of the KL divergence minimization, the regression NN maps latent vector \( z_k \) to \( \hat{y} \) by the mapping, \( \hat{y} = g_k(z_k | y) \), which consists of \( P \) layers, \( g_k^{(1)}, \ldots, g_k^{(P)} \). The weight \( W_k^{(P)} \) and bias \( b_k^{(P)} \) of the last layer estimate the output \( \hat{y} = [\hat{y}_1, \hat{y}_2] \), which is represented in the normalized Cartesian coordinate in our case:

\[
\hat{y} = \phi_k^{(P)} = W_k^{(P)} \phi_k^{(P-1)} + b_k^{(P)}.
\]

Then its regression loss is evaluated in terms of Euclidean distance with the known true \( y \) for supervised learning. We jointly minimize the KL divergence and the regression loss by updating the weight and bias parameters.

In order to achieve supervised view-oriented learning, we utilize the given view label \( u \) during the training. For every multiview training sample, the mappings \( h_k(x_k) \) and \( g_k(z_k | y) \) are optimized to
generate $z_k$ only for the informative views as follows:

$$
\begin{align*}
\begin{cases}
  \min_{w_k,b_k} W_k, b_k \{ (y_1 - \hat{y}_1)^2 + (y_2 - \hat{y}_2)^2 \} + KL[p(z_k) || p(z_k)] & \text{if } u_k = 1, \\
  \text{Do nothing} & \text{if } u_k = 0.
\end{cases}
\end{align*}
$$

We expect the NNs properly extract latent features with excluding the non-informative data view. Although trained weights and biases can generate the latent feature vector $z_k$ for every test sample regardless of the condition of $u_k$, the invisible latent features from two different samples may have a strong correlation if both are informative in a certain view $k$ (i.e., $u_k = 1$). Now $z$ for every training and test sample becomes a new input in the next step. In the following section 4.2, we will introduce an additional network to enhance regression using the latent features and their invisible correlation.

### 4.2 View-Classified Regression Network

The view-classified regression network, as described in the right grey box in Figure 2, uses an intermediate reweighting method $\alpha$ where

$$
\alpha = \frac{u_k}{\sum_{i=1}^{K} u_i}.
$$

The view classification NN of layers $h_Q^{(1)}, \ldots, h_Q^{(Q)}$ maps $z$ to $\hat{u}$ such that $\hat{u} = h_Q(z|\hat{u})$. Starting from the first layer input $z (\phi_Q^{(0)} = z)$, we calculate the view classification result $\hat{u}$ from the output layer through the softmax activation:

$$
\hat{u} = \text{softmax}(W_Q^{(Q)} \phi_Q^{(Q-1)} + b_Q^{(Q)}),
$$

where $W_Q^{(q)}$ and $b_Q^{(q)}$ denote the weight and bias of the layer $q \in \{1, \ldots, Q\}$, respectively.

Regarding $\hat{u}$ as the reweight parameter, we recompute each subset $z_k$ to $z'_k$:

$$
z'_k = \hat{u}_k \odot z_k,
$$

and use it as a regression input. With concatenated $z' = \{z'_1, \ldots, z'_K\}$, the regression NN of layers $h_R^{(1)}, \ldots, h_R^{(R)}$ maps $z'$ to $\hat{y}$, such that $\hat{y} = h_R(z'|y)$, and obtains the Cartesian coordinate output $\hat{y} = [\hat{y}_1, \hat{y}_2]$ in our case:

$$
\hat{y} = \phi_R^{(R)} = W_R^{(R)} \phi_R^{(R-1)} + b_R^{(R)},
$$

where $W_R^{(r)}$ and $b_R^{(r)}$ are the weight and bias of the layer $r \in \{1, \ldots, R\}$. The NNs update parameters $W_Q, b_Q, W_R$, and $b_R$ to jointly minimize both Euclidean losses:

$$
\min_{w_Q, b_Q, w_R, b_R} \alpha\{(y_1 - \hat{y}_1)^2 + (y_2 - \hat{y}_2)^2\} + (1 - \alpha)\{\sum_{k}(\hat{u}_k - \hat{u}_k)^2\},
$$

where $\alpha \in (0, 1)$ denotes a trade-off parameter between two losses. With the fingerprint database consisting of weights and biases, for the test data, we can obtain the regression output $(\hat{y}_1, \hat{y}_2)$ which is the localization result. Strictly speaking, our reweighting is different from the existing iterative reweight (IR) methods [30, 31], which sought the reweighting based on their gradient directions. In contrast, we suggest a simpler method that derives reweight parameters based on the given information. Our reweighting method assists the supervised learning system to decide which data view should be considered more importantly and further to achieve better performance.
5 Field Experiment

We apply the VSDL system for indoor localization on a two-corridor real building environment with 43 training points and 9 test points, as in Figure 3a. Each corridor is 7 m long, where training and test points are spread by 0.5 m spacing. We install seven APs consisting of 3-antenna IWL 5300 NIC in laptop computers, which are placed at the corners of the corridors. For a transmitter, the same laptop with a single antenna is used to transmit WiFi packets on channel 36 at 5.18 GHz. The APs receive packets from the transmitter at the same time using the monitor mode and we combine them as a multiview input at the server side. Each AP receives the WiFi packet with three antennas to form three Tx-Rx radio channels. Each channel produces a CSI vector consisting of 30 subcarrier CSIs ($I = 30$). We then take one of the three CSI vectors as the reference to produce two relative CSI vectors. In this way, we obtain an input sample $x$ consisting of $420 = 7 \times (3 - 1) \times 30$ relative CSIs. We collect 100 sample packets for every training and test points in a noisy environment, which manifests high data fluctuation that makes it difficult to estimate the location by other analytical methods.

![Figure 3: Two-corridor experiment. (a) Localization topology with 43 training points (open circle) and 9 test points (solid circle). There are seven APs to receive packets from the points at the same time. (b) Regression results from a variational DL system with no exclusion of non-informative views. (c) Regression results from the VSDL system with exclusion of non-informative views, which finds the dominant views effectively to enhance localization accuracy.](image)

In this scenario, we divide input $x$ into two views $x_1$ and $x_2$ ($K = 2$), which represent AP associations with corridors 1 and 2, respectively. Therefore, $x_1$ has information of AP1 to AP5 and $x_2$ has that of AP3 to AP7. Along with the CSI data, the location label $y$ in the Cartesian coordinate and the given corridor label $u$ are utilized for training. There are three cases for $u$ depending on the training location: 1) only to corridor 1 ($u = [1, 0]$), 2) only to corridor 2 ($u = [0, 1]$), and 3) to both corridors 1 and 2 ($u = [1, 1]$). The NNs for view-oriented learning update parameters only when the view is informative ($u_k = 1$). Here, the information from AP3 to AP5 are common in both views and considered to be informative for every sample.

![Figure 4: Regression and view classification losses versus $J$.](image)  
![Figure 5: Regression and view classification losses versus $\alpha$.](image)

Along the entire system, we set the number of hidden layers $L, P, Q,$ and $R$ to be three. The number of nodes of each layer decreases from 1000 to 500. For the feature output of every layer except for the last one, the Relu activation is used. Adam optimizers are used to update the parameters with learning rate of $10^{-5}$. First, we aim to verify if our joint optimization works well. As seen in
Figure 4: the losses of regression and view classification are minimized together for any given number of latent variable $J$. This trend implies that our system properly operates to obtain an advanced regression result assisted by the view selection. In addition, we obtain the best regression result with $J = 120$ rather than keeping the size of the input CSI vector, which corresponds to a mean-field feature compression ratio of $2/7$. There should be a sufficiently large number of variables to estimate the posterior of the CSI data, while too many variables cause overfitting of the network. Obviously, $J$ depends on the scenario as well as the application requirement.

In our VSDL system, the trade-off parameter $\alpha$ can influence the performance of regression as seen in Figure 5. As $\alpha$ approaches 0, the network becomes too sensitive on view classification, and it makes both losses worse. On the other hand, as $\alpha$ approaches 1, high view classification loss occurs, resulting in poor regression (localization) accuracy. We obtain the best regression accuracy with $\alpha = 0.5$.

Figures 3b and 3c show comparisons of the regression results from a simple variational DL and our VSDL with $J = 120$ and $\alpha = 0.5$. We take three test locations as the representative cases. The locations A and B are near the end of corridors and location C is in the intersection of the two corridors. The test results for the location A, B, and C are shown in blue, red, and green, respectively. In terms of multiview data learning, the variational DL extracts features from all corridor views including non-informative views. Therefore, as seen in Figure 3b, the regression results for many cases are located somehow outside the topology of training. In contrast, the VSDL system updates the learning parameters only for informative views to classify the dominant view and hence to achieve better localization, as seen in Figure 3c.

Figure 6: Localization error CDF of the systems. The proposed VSDL significantly outperforms other previous systems.

| Algorithm     | Localization error (m) |
|---------------|------------------------|
| VSDL (proposed) | 0.7715                 |
| VDL (variational DL) | 1.0607             |
| SVR           | 1.1037                 |
| DNN           | 1.1246                 |
| BiLoc         | 1.1844                 |
| CiFi          | 1.9739                 |

Further, we compare the proposed VSDL with several existing machine learning systems. To discriminate the CSI data, both classification and regression methods were introduced in previous papers to improve the localization accuracy. In our experiment environment, as well as simple variational DL, we implemented RBM based classification BiLoc [14], CNN based classification CiFi [17], SVM based regression SVR [19], and stand-alone DNN based regression. Figure 6 and Table 1 show the comparison results. We do not plot the results of CiFi in this scenario, since the convolutional analysis from batch information cannot extract proper features and results in a very poor localization accuracy of 1.97 m. First, the variational DL, whose results are described in 3b, outperforms other existing systems due to the usage of variational inference. Here, we observe that the introduction of the variational inference brings the key advantage for WiFi CSI localization in a noisy radio channel. In addition, the VSDL system with novel two-stage view-selective learning on the variational inference base significantly improves the localization accuracy by 30%, from 1.10 m to 0.77 m. As the VSDL is very scalable by the nature of its design, we expect further performance improvement in environments with more corridors.

6 Conclusions

WiFi device localization has been a very attractive area of study as WiFi networks are omnipresent to provide network application services to anonymous users in these days, anticipated to open new business opportunities as well as new technical challenges. The technical performance of WiFi localization has been improved disruptively by the use of channel state information measured at multiple receive APs.
In this paper, we introduce a machine learning design that combines the variational deep learning very effectively in multiview learning architecture. We report an observation that the latent vectors generated at the intermediate layer of variational deep learning form strong feature behaviors to provide classification of effective view selection that enhance the accuracy of localization by a great deal. Our system, the view-selective deep learning, or the VSDL, achieves a localization accuracy of 0.77 m, which manifests a more than 30 % improvement in a two-corridor field experiment compared with the best known system based on SVM. The VSDL is completely scalable as exploiting the benefit of multiview-based regression, and hence the WiFi localization network can be expanded with no limit, such as in complex building structure. Our design of extracting features in the latent space to deal with informative and non-informative views in a multiview variational deep learning network is very powerful so that it can be applied to various applications with no limit on scalability.

Broader Impact

The indoor localization with radio signals, for example, WiFi radio signals, which finds the location of a mobile device very accurately can create a great deal of impact in mobile service application. This can be a benefit to off-line stores and services such as in a shopping mall, hospital, and public buildings, where location-based services can directly improve quality of experiences, especially when associated with social network services. Of course, such network features of localization can also harm the privacy of people in public. Radio localization may fail in a hot spot in the sense of crowded radio network traffic. However, such a failure may not cause critical problems except for some frustration with internet-based applications.

References

[1] Matthew D Hoffman, David M Blei, Chong Wang, and John Paisley. Stochastic variational inference. The Journal of Machine Learning Research, 14(1):1303–1347, 2013.

[2] Bo Zhao, Xiao Wu, Zhi-Qi Cheng, Hao Liu, Zequn Jie, and Jiashi Feng. Multi-view image generation from a single-view. In Proceedings of the 26th ACM international conference on Multimedia, pages 383–391, 2018.

[3] Nat Dilokthanakul, Pedro AM Mediano, Marta Garnelo, Matthew CH Lee, Hugh Salimbeni, Kai Arulkumar, and Murray Shanahan. Deep unsupervised clustering with gaussian mixture variational autoencoders. arXiv preprint arXiv:1611.02648, 2016.

[4] Weidi Xu, Haoze Sun, Chao Deng, and Ying Tan. Variational autoencoder for semi-supervised text classification. In Thirty-First AAAI Conference on Artificial Intelligence, 2017.

[5] Mark Girolami and Simon Rogers. Variational bayesian multinomial probit regression with gaussian process priors. Neural Computation, 18(8):1790–1817, 2006.

[6] Andrew G Wilson, Zhitong Hu, Russ R Salakhutdinov, and Eric P Xing. Stochastic variational deep kernel learning. In Advances in Neural Information Processing Systems, pages 2586–2594, 2016.

[7] Martin J Wainwright and Michael I Jordan. Graphical models, exponential families, and variational inference. Foundations and Trends® in Machine Learning, 1(1-2):1–305, 2008.

[8] Shubham Tulsiani, Tinghui Zhou, Alexei A Efros, and Jitendra Malik. Multi-view supervision for single-view reconstruction via differentiable ray consistency. In Proceedings of the IEEE conference on computer vision and pattern recognition, pages 2626–2634, 2017.

[9] Guoquan Li, Enxu Geng, Zhouyang Ye, Yongjun Xu, Jinhao Lin, and Yu Pang. Indoor positioning algorithm based on the improved rssi distance model. Sensors, 18(9):2820, 2018.

[10] Jie Xiong and Kyle Jamieson. Arraytrack: A fine-grained indoor location system. In Presented as part of the 10th {USENIX} Symposium on Networked Systems Design and Implementation ({NSDI} ’13), pages 71–84, 2013.

[11] Manikanta Kotaru, Kiran Joshi, Dinesh Bharadia, and Sachin Katti. Spotfi: Decimeter level localization using wifi. In Proceedings of the 2015 ACM Conference on Special Interest Group on Data Communication, pages 269–282, 2015.
[12] Xuyu Wang, Lingjun Gao, Shiwen Mao, and Santosh Pandey. Deepfi: Deep learning for indoor fingerprinting using channel state information. In 2015 IEEE wireless communications and networking conference (WCNC), pages 1666–1671. IEEE, 2015.

[13] Xiandi Li, Jingshi Shi, and Jianli Zhao. Defe: indoor localization based on channel state information feature using deep learning. In Journal of Physics: Conference Series, volume 1303, page 012067. IOP Publishing, 2019.

[14] Xuyu Wang, Lingjun Gao, and Shiwen Mao. Biloc: Bi-modal deep learning for indoor localization with commodity 5ghz wifi. IEEE Access, 5:4209–4220, 2017.

[15] Brieuc Berruet, Oumaya Baala, Alexandre Caminada, and Valery Guillet. Delfin: A deep learning based csi fingerprinting indoor localization in iot context. In 2018 International Conference on Indoor Positioning and Indoor Navigation (IPIN), pages 1–8. IEEE, 2018.

[16] Hao Chen, Yifan Zhang, Wei Li, Xiaofeng Tao, and Ping Zhang. Confi: Convolutional neural networks based indoor wi-fi localization using channel state information. IEEE Access, 5:18066–18074, 2017.

[17] Xuyu Wang, Xiangyu Wang, and Shiwen Mao. Cifi: Deep convolutional neural networks for indoor localization with 5 ghz wi-fi. In 2017 IEEE International Conference on Communications (ICC), pages 1–6. IEEE, 2017.

[18] Pengpeng Chen, Fen Liu, Shouwan Gao, Peihao Li, Xu Yang, and Qiang Niu. Smartphone-based indoor fingerprinting localization using channel state information. IEEE Access, 7:180609–180619, 2019.

[19] Rui Zhou, Jiesong Chen, Xiang Lu, and Jia Wu. Csi fingerprinting with svm regression to achieve device-free passive localization. In 2017 IEEE 18th International Symposium on A World of Wireless, Mobile and Multimedia Networks (WoWMoM), pages 1–9. IEEE, 2017.

[20] Zhihui Gao, Yunfan Gao, Sulei Wang, Dan Li, Yuedong Xu, and Hongbo Jiang. Crisloc: Reconstructable csi fingerprinting for indoor smartphone localization. arXiv preprint arXiv:1910.06895, 2019.

[21] Hsiao-Chien Tsai, Chun-Jie Chiu, Po-Hsuan Tseng, and Kai-Ten Feng. Refined autoencoder-based csi hidden feature extraction for indoor spot localization. In 2018 IEEE 88th vehicular technology conference (VTC-Fall), pages 1–5. IEEE, 2018.

[22] Xiaochao Dang, Jiaju Ren, Zhanjun Hao, Yili Hei, Xuhao Tang, and Yan Yan. A novel indoor localization method using passive phase difference fingerprinting based on channel state information. International Journal of Distributed Sensor Networks, 15(4):1550147719844099, 2019.

[23] Chaur-Heh Hsieh, Jen-Yang Chen, and Bo-Hong Nien. Deep learning-based indoor localization using received signal strength and channel state information. IEEE access, 7:33256–33267, 2019.

[24] Tahsina Farah Sanam and Hana Godrich. A multi-view discriminant learning approach for indoor localization using amplitude and phase features of csi. IEEE Access, 8:59947–59959, 2020.

[25] Daniel Halperin, Wenjun Hu, Anmol Sheth, and David Wetherall. Predictable 802.11 packet delivery from wireless channel measurements. ACM SIGCOMM Computer Communication Review, 41(4):159–170, 2011.

[26] Ralph Schmidt. Multiple emitter location and signal parameter estimation. IEEE transactions on antennas and propagation, 34(3):276–280, 1986.

[27] Asaf Tzur, Ofer Amrani, and Avishai Wool. Direction finding of rogue wi-fi access points using an off-the-shelf mimo–ofdm receiver. Physical Communication, 17:149–164, 2015.

[28] Ion Muslea, Steven Minton, and Craig A Knoblock. Selective sampling with redundant views. In AAAI/IAAI, pages 621–626, 2000.

[29] William Stafford Noble et al. Support vector machine applications in computational biology. Kernel methods in computational biology, 71:92, 2004.

[30] Rick Chartrand and Wotao Yin. Iteratively reweighted algorithms for compressive sensing. In 2008 IEEE International Conference on Acoustics, Speech and Signal Processing, pages 3869–3872. IEEE, 2008.

[31] Karthik Mohan and Maryam Fazel. Iterative reweighted algorithms for matrix rank minimization. Journal of Machine Learning Research, 13(Nov):3441–3473, 2012.