UNIVERSITÀ DI PISA

Dipartimento di Informatica

Corso di Laurea Triennale in Informatica

Localizzazione Indoor Basata su Beacon Bluetooth a Bassa Potenza Attraverso Tecniche di Deep Learning

un progetto realizzato per Consorzio Metis e ASL Toscana

Relatore:
Prof. GianLuigi Ferrari

Presentata da:
Marco Pampaloni

Anno Accademico 2019/2020

Sommario

Il problema della Localizzazione Indoor si è rivelato di particolare interesse pratico negli ultimi anni. Questa tesi mostra come moderne tecniche di Deep Learning possano risultare determinanti nella corretta risoluzione di tale problema.

L'approccio analizzato sfrutta una rete neurale convoluzionale (CNN) profonda: l'input del modello è caratterizzato da una serie temporale di segnali broadcast *Bluetooth Low Energy* (BLE) emessi da un insieme di Beacon disposti all'interno dell'edificio adibito alla Localizzazione Indoor, mentre l'output è una coppia di coordinate relative alla posizione all'interno dell'edificio stesso. Sono state inoltre utilizzate varie tecniche di *data augmentation* per produrre un dataset di grandi dimensioni sulla base dei campionamenti dei segnali effettuati in loco.

A seguito dell'addestramento, il modello utilizzato ha mostrato un errore medio assoluto (MAE) sul dataset di test pari a *30cm*, esibendo una discreta affidabilità anche rispetto a variazioni significative dei segnali dovute al rumore ambientale. Un ensemble di modelli, ognuno addestrato con diversi iperparametri, ha permesso di ridurre l'errore medio fino a circa *26cm*.

Il modello prodotto risulta eseguibile in tempo reale su dispositivi mobile con ridotte capacità computazionali, rendendolo particolarmente adatto alla cosiddetta navigazione *"blue-dot"* all'interno di contesti Indoor. Tuttavia si evidenzia come la variazione dell'output del modello possa risultare in una navigazione poco fluida. Per arginare questo problema viene applicato un filtro di Kalman al modello e viene sfruttato il sensore inerziale dello smartphone per produrre un'euristica utile a individuare i movimenti dell'utente.

Indice

Capitolo 1

Introduzione

1.1 Localizzazione Indoor

Il problema della Localizzazione Indoor consiste nell'individuazione di un utente all'interno di uno spazio chiuso e in riferimento a un sistema di coordinate predefinito. Tale sistema di coordinate, relativo ad un determinato edificio, può essere poi espresso in termini georeferenziali conoscendo la precisa dislocazione geografica del locale in questione.
La localizzazione indoor apre le porte a diverse possibilità nel campo dell'esperienza utente all'interno di edifici pubblici, nel settore della gestione dei flussi di persone, della sicurezza e della contingentazione. Attraverso l'impiego di tale tecnologia è possibile coadiuvare la navigazione degli utenti all'interno di edifici complessi e migliorare l'esperienza individuale di persone affette da disabilità. Per ottenere questi risultati è però richiesto un certo grado di precisione, di affidabilità, di efficienza e di sicurezza nella gestione della privacy dei dati di localizzazione degli utenti. Inoltre la tecnologia scelta per risolvere il problema, per essere fruibile, deve avere come ulteriore requisito il basso impatto economico.

1.2 Soluzioni Tecnologiche

Nel corso degli anni sono state implementati diversi sistemi di localizzazione indoor, che possiamo dividere in due macrocategorie: soluzioni ad-hoc e soluzioni che sfruttano tec-

nologie esistenti. Nel primo caso si fornisce all'utente l'attrezzatura necessaria ad essere localizzato, mentre nel secondo si utilizza un dispositivo mobile di proprietà dell'utilizzatore. Spesso tale dispositivo è uno smartphone.
I sistemi che implementano tecnologie sviluppate ad-hoc, sono spesso più efficienti, più precisi e flessibili. Tuttavia, il loro impiego rimane limitato a causa dell'alto costo di progettazione, installazione e di gestione. È poi richiesto che ad ogni utente che intenda essere localizzato sia assegnato un dispositivo che si interfacci col sistema impiegato.
Per l'impiego su larga scala, un sistema di localizzazione indoor deve essere facilmente utilizzabile dalle masse e non deve richiedere particolari requisiti tecnologici.

TODO: Inserire riferimenti bibliografici che mettano in comparazione le varie tecnologie utilizzate, ad-hoc e non, e in particolare mostrino i risultati dei sistemi che sfruttano il Bluetooth.

1.3 Bluetooth Low Energy

La tecnologia *Bluetooth* è talmente pervasiva che ogni smartphone in circolazione ne implementa il protocollo, mostrandosi particolarmente adeguata alla risoluzione del problema in esame. Nello specifico, *Bluetooth Low Energy* (BLE) è un protocollo che riduce notevolmente il consumo energetico dei dispositivi che ne sfruttano le capacità.
La soluzione riportata in questo documento prevede l'utilizzo di una serie di beacon BLE programmabili, ciascuno installato in un punto significativo dell'edificio e configurato per emettere un segnale broadcast con una frequenza di circa 50Hz. La potenza dei segnali viene quindi utilizzata per produrre, attraverso l'utilizzo di una rete neurale artificiale, una coppia di coordinate rappresentative della posizione dell'utente all'interno dell'edificio. Ciò viene reso possibile da una fase preliminare in cui viene mappata la superficie del locale raccogliendo i segnali ricevuti dai beacon in vari punti. Per ogni punto della superficie mappato si registra una serie temporale di segnali, dei quali si considera solo il valore *RSSI*, ovvero la potenza del segnale nel punto in cui questo viene ricevuto.

Il modello utilizzato è di fatto completamente agnostico rispetto all'ubicazione dei beacon installati, fin quando questa sia unica e non mutata nel tempo.
L'utilizzo di tale sistema assicura il completo anonimato dell'utente, il quale non necessita di condividere la propria posizione, essendo quest'ultima calcolata direttamente sul suo smartphone in funzione dei segnali che riceve.

Questa tesi si pone l'obiettivo di descrivere nello specifico la rete neurale progettata per risolvere il problema, le tecniche utilizzate per alzare il grado di precisione del modello e le principali differenze rispetto a modelli già esistenti.

TODO: introdurre breve descrizione dei capitoli

1.4 RSSI e propagazione del segnale

1.5 Variabilità e rumore di fondo: requisiti di usabilità

1.6 Installazione dei Beacon e Acquisizione dei Dati

Capitolo 2

Deep Learning

In questo capitolo saranno introdotti i concetti fondamentali alla base delle moderne tecniche di Deep Learning e le strutture matematiche necessarie alla loro comprensione.

2.1 Machine Learning

Il *Machine Learning*, o apprendimento automatico, è un insieme di tecniche e algoritmi che consente a dei programmi di "imparare" a svolgere un determinato compito sulla base di esperienze pregresse, senza bisogno da parte del programmatore di specificare come eseguire tali mansioni.

2.1.1 Regressione Lineare

Un classico esempio di algoritmo di machine learning è quello della regressione lineare. Scopo dell'algoritmo è predire l'output di una determinata funzione. Si consideri quindi un vettore $\boldsymbol{x} \in \mathbb{R}^n$, e un valore scalare $\hat{y} = \boldsymbol{\theta}^\intercal \boldsymbol{x}$. Il vettore $\boldsymbol{\theta}$ introduce i parametri del modello, mentre $\hat{y}$ ne rappresenta l'output, che è una funzione lineare di $\boldsymbol{x}$. Siano quindi $\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_m$ dei vettori in $\mathbb{R}^n$ e $y_1, y_2, \ldots, y_m$ i corrispettivi valori della funzione $f(\boldsymbol{x}_i)$ che stiamo cercando di approssimare.

Perchè la funzione $\hat{y}(\boldsymbol{x})$ approssimi $f(\boldsymbol{x})$ è necessario che i parametri $\boldsymbol{\theta}$ del modello si adattino in modo da minimizzare la differenza tra l'output prodotto dal modello e la cosiddetta *ground truth*: $y = f(\boldsymbol{x})$. A questo scopo si definisce una metrica di errore propria del processo di apprendimento: l'errore quadratico medio (MSE dall'inglese)

$$MSE = \frac{1}{m} \sum_{i=1}^{m} (\hat{y} - y_i)^2 \tag{2.1}$$

Per minimizzare l'errore sul nostro dataset di test è sufficiente porre a zero la derivata, rispetto a $\boldsymbol{\theta}$, della nostra funzione di costo. Nel caso della regressione lineare è possibile risolvere l'equazione risultante ottenendo un sistema di equazioni che prende il nome di *normal equations*. Esistono tuttavia metodi numerici iterativi che si basano sulle informazioni fornite dal gradiente della funzione di costo che permettono di aggiornare i parametri del modello cercando di ridurne l'errore, anche nel caso di modelli non lineari. L'algoritmo su cui si basano molti dei moderni metodi di apprendimento del Deep Learning è il *gradient descent* o metodo del gradiente.

Il metodo del gradiente aggiorna i parametri $\boldsymbol{\theta}$ del modello secondo la seguente regola:

$$\boldsymbol{\theta}' = \boldsymbol{\theta} - \eta \nabla_{\boldsymbol{\theta}} J(\boldsymbol{\theta}) \tag{2.2}$$

dove $J(\boldsymbol{\theta})$ rappresenta la funzione di costo associata al modello, mentre η è un coefficiente chiamato *learning rate*. Un'interpretazione dell'algoritmo è data dalle informazioni sulla monotonia ottenute dal gradiente della funzione di costo: per η abbastanza piccolo risulta $J(\boldsymbol{\theta}') \leq J(\boldsymbol{\theta})$ poichè il gradiente negativo di J determina la direzione di massima decrescita della funzione[2]. Ne consegue che applicando ricorsivamente la regola di aggiornamento del metodo del gradiente, la nostra funzione $\hat{y}$ tenderà ad avvicinarsi alla funzione originale y, minimizzando la funzione di costo. Possiamo intepretare il coefficiente η come la velocità con cui seguire la pendenza della funzione di errore.

TODO: Inserire un'immagine che fornisca un'intuizione delfunzionamentodel gradient descent

Nel caso di un modello di regressione lineare che usa l'MSE come funzione di costo, risulta:

$$\begin{aligned}
&\forall j \in 1, \dots, n : \\
\frac{\partial}{\partial \theta_j} MSE &= \frac{\partial}{\partial \theta_j} \frac{1}{m} \sum_{i=1}^{m} (\hat{y}_i - y_i)^2 \\
&= \frac{\partial}{\partial \theta_j} \frac{1}{m} \sum_{i=1}^{m} (\boldsymbol{\theta}^\intercal \boldsymbol{x}_i - y_i)^2 \\
&= \frac{1}{m} \sum_{i=1}^{m} \frac{\partial}{\partial \theta_j} (\boldsymbol{\theta}^\intercal \boldsymbol{x}_i - y_i)^2 \\
&= \frac{2}{m} \sum_{i=1}^{m} (\boldsymbol{\theta}^\intercal \boldsymbol{x}_i - y_i) x_i^{(j)}
\end{aligned}$$

Ovvero abbiamo che la regola di aggiornamento è:

$$\theta_j' = \theta_j - \eta \frac{\partial}{\partial \theta_j} J(\boldsymbol{\theta}) = \theta_j - \eta (\frac{2}{m} \sum_{i=1}^{m} (\boldsymbol{\theta}^\intercal \boldsymbol{x}_i - y_i) x_i^{(j)}), \quad \forall j \in \{1, \dots, n\}$$

2.1.2 Perceptron

Il *Perceptron* è il modello che ha posto le basi per le moderne reti neurali artificiali e il deep learning. Esso prende spunto dalla neurologia, cercando di imitare il comportamento dei neuroni del cervello umano, con ovvie limitazioni e senza presunzione di volerne fornire una simulazione accurata del funzionamento. Una schematizzazione del modello è descritta in Figura 2.1

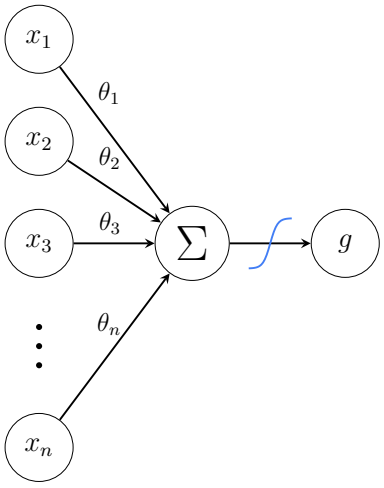


Figura 2.1: Schematizzazione del Perceptron

Il perceptron è molto simile al modello di regressione lineare descritto precedentemente, ma si distingue per un fattore fondamentale: la non linearità. L'output del perceptron è infatti definito come:

$$\hat{y}(\boldsymbol{x}) = g(\boldsymbol{\theta}^\intercal \boldsymbol{x})$$

dove g è una funzione *sigmoidea*, cioè una funzione che ha un andamento a "S", con due asintoti orizzontali come in Figura 2.2: di solito è utilizzata la funzione logistica $g(x) = \dfrac{1}{1+e^{-x}}$.

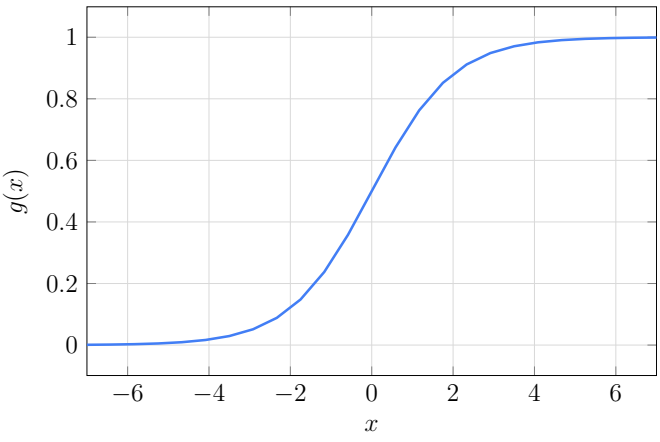


Figura 2.2: Funzione sigmoidea $g(x) = \dfrac{1}{1+e^{-x}}$

Malgrado la nonlinearità del modello, il perceptron non è in grado di approssimare molte classi di funzioni. È famoso l'esempio della funzione XOR, la quale non può essere imparata dal perceptron. Per questo motivo è stata sviluppata un'estensione del modello che prende il nome di *Multi Layer Perceptron*.

2.2 Multi Layer Perceptron

Il Multi Layer Percepton (MLP) è uno dei modelli più emblematici del Deep Learning. Esso si basa sui concetti descritti finora ed è la naturale estensione del perceptron. L'MLP risolve infatti il principale problema del modello su cui è fondato, essendo in grado di approssimare qualsiasi tipo di funzione continua con precisione arbitraria, sotto l'assunzione che la funzione di attivazione sia non polinomiale[3].

L'MLP è un modello di machine learning che fa parte della categoria delle reti neurali, in quanto è composto da un insieme di neuroni artificiali (perceptron) disposti su più livelli e interconnessi tra di loro. Una schematizzazione del modello è fornita in Figura 2.3.

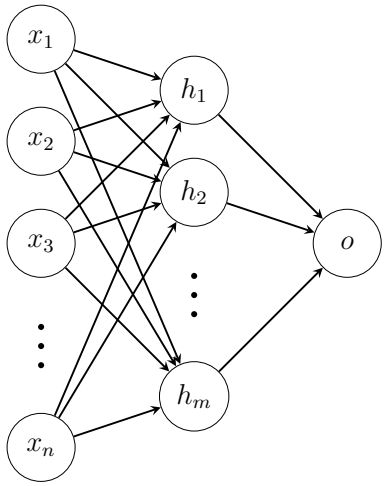


Figura 2.3: Schematizzazione del Multi Layer Perceptron: in Figura è mostrata una rete neurale con n valori di input, un singolo hidden layer con m neuroni e un solo output. Si noti che non si è limitati ad usare un solo neurone di output. Sono omessi per chiarezza i parametri del modello e le funzioni di attivazione, riassunte in questo caso all'interno di ogni neurone.

Nello specifico, l'MLP presenta un primo livello N-dimensionale che coincide con il

suo input, un livello di output M-dimensionale e una serie arbitraria di livelli intermedi (*hidden layers*) di ampiezza variabile. Ogni nodo del livello precedente è connesso con ogni neurone del livello successivo e ogni neurone del modello si comporta come un singolo perceptron, ovvero esegue una somma pesata degli input ricevuti dal livello precedente e produce in output il risultato attraverso una funzione di attivazione non-lineare. Per questo motivo i parametri del modello, cioè i coefficienti con cui vengono sommati gli input di ogni nodo, sono uno per ogni connessione. Questo tipo di modelli vengono anche chiamati *Feed Forward Networks*, in quanto le connessioni tra nodi sono dirette soltanto verso i livelli direttamente successivi, e mai il contrario.

Sia $\boldsymbol{\Theta}^{\ell}$ la matrice dei pesi delle connessioni tra il livello $\ell - 1$ e quello successivo, definita in modo che Θ_{ij}^{ℓ} rappresenti il coefficiente relativo alla connessione tra il neurone i-esimo del livello ℓ e il neurone j-esimo del livello $\ell - 1$. Questa configurazione della matrice dei coefficienti, nonostante sia poco intuitiva, permette di esprimere l'insieme di valori uscenti da un generico livello ℓ sotto forma di prodotto: $\boldsymbol{h}^{\ell} = \boldsymbol{\Theta}^{\ell}\boldsymbol{h}^{\ell-1}$. Se L è il numero di livelli intermedi della rete neurale, risulta:

$$\begin{aligned} o(\boldsymbol{x}) &= \boldsymbol{\Theta}^{L+1}\boldsymbol{h}^{L} \\ &= \boldsymbol{\Theta}^{L+1} g(\boldsymbol{\Theta}^{L}\boldsymbol{h}^{L-1}) \\ &= \boldsymbol{\Theta}^{L+1} g(\boldsymbol{\Theta}^{L} g(\dots g(\boldsymbol{\Theta}^{1}\boldsymbol{x}))) \end{aligned}$$

2.3 BackPropagation

Per utilizzare il metodo del gradiente con una rete neurale Feed Forward, è necessario calcolare la derivata della funzione di costo del modello. Tuttavia nell'MLP essa dipende sia dai parametri del livello immediatamente precedente che da tutti i parametri dei livelli più bassi, per ricorsione. Si può pensare al BackPropagation come un modo per propagare all'indietro le informazioni relative all'errore commesso dai nodi di output rispetto ai target del dataset di addestramento.

L'algoritmo calcola il gradiente della funzione di costo in modo automatico applicando ricorsivamente lungo il grafo di computazione di $J(\boldsymbol{\Theta})$ la "regola della catena", ovvero la regola di derivazione per le funzioni composte: $(f \circ g)' = (f' \circ g) \cdot g'$. Esso sfrutta la tecnica della programmazione dinamica per evitare di ricalcolare più volte la derivata di branch comuni dell'albero di computazione, rendendolo di fatto un metodo molto efficiente.

2.4 Attivazione: ReLU

Come descritto, la funzione di attivazione conferisce al modello la nonlinearità. Ciò è fondamentale perchè l'MLP riesca ad approssimare funzioni continue arbitrarie. Se non ci fosse alcuna funzione di attivazione o se questa fosse lineare rispetto ai parametri del modello, l'output della rete si potrebbe esprimere come una semplice applicazione lineare, rendendolo di fatto equivalente a un semplice modello di regressione lineare.

Con l'avvento delle reti neurali profonde e di nuove architetture più complesse, l'utilizzo delle funzioni di attivazioni logistiche è andato calando. Una delle cause è il cosiddetto "vanishing gradient", un problema che insorge nel calcolare numericamente il gradiente della funzione di costo di una rete neurale con molti livelli. In questo caso l'uso della funzione sigmoidea è in generale sconsigliato poichè il suo valore "satura" a causa dei suoi asintoti e questo rende la sua derivata prossima a zero. Il metodo del gradiente risulta quindi poco efficace nell'aggiornare i parametri del modello. Per questo motivo è stata introdotta la funzione ReLU (dall'inglese *rectified linear unit*) definita nel modo seguente: $g(x) = \max\{0, x\}$. L'operazione di rettificazione è mostrata in Figura 2.4.

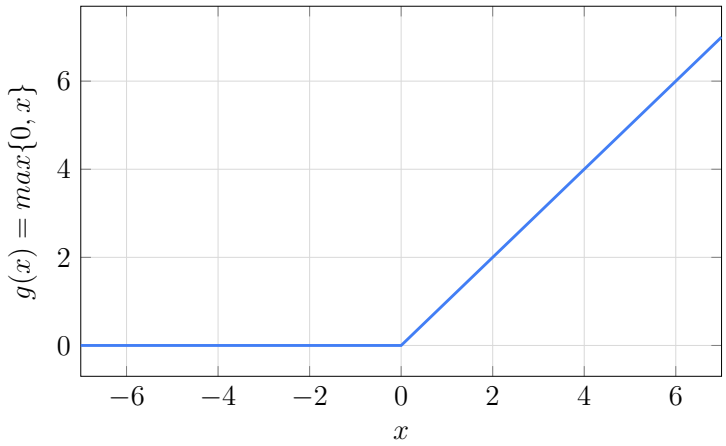


Figura 2.4: La funzione di attivazione ReLU

L'output del rettificatore, essendo esso quasi una funzione lineare, mantiene le informazioni del gradiente in modo migliore rispetto al sigmoide, rendendo la convergenza del metodo del gradiente più veloce.

2.5 Reti Neurali Convoluzionali

Una particolare categoria di reti feed forward è quella delle reti neurali convoluzionali, o *convolutional neural networks* (CNN) in inglese. Una CNN non è altro che una rete neurale che adotta l'operazione di *convoluzione* in almeno uno dei suoi layer, al posto della più comune moltiplicazione matriciale.

L'operazione di convoluzione, usualmente indicata con il simbolo $*$, è definita per due funzioni continue f e g come:

$$(f * g)(t) = \int_{-\infty}^{\infty} f(x)k(t-x)dx$$

mentre, nel caso discreto, è invece definita nel seguente modo:

$$(f * g)(t) = \sum_{x=-\infty}^{\infty} f(x)k(t-x)$$

Poichè in una rete neurale le funzioni f e g sono rappresentate generalmente da tensori, l'operazione può essere limitata ai soli elementi di tali insiemi di valori, con l'assunzione che le due funzioni siano nulle ovunque tranne nei punti in cui sono definiti i tensori. Nel caso di input bidimensionali, come è quello delle immagini, si ha quindi:

$$\begin{aligned}(X * K)(i,j) &= \sum_m \sum_n X(m,n)K(i-m, j-n) \\ &= \sum_m \sum_n X(i-m, j-n)K(m,n)\end{aligned}$$

in cui X rappresenta l'input dell'operazione, mentre K è chiamato *kernel*, o filtro e costituisce i parametri adattivi della CNN. Nella pratica è tuttavia più comune utilizzare l'operazione di *cross-correlation*, definita come

$$(K * X)(i,j) \sum_m \sum_n X(i+m, j+n)K(m,n)$$

Tale operazione equivale a spostare il filtro K lungo le due dimensioni dell'input X, eseguendo una moltiplicazione elemento per elemento tra i due tensori e sommandone i risultati. Un esempio è illustrato in Figura 2.5.

2.5.1 Pooling

L'output della convoluzione è chiamato *feature map* e generalmente ne vengono calcolate svariate decine per ogni livello, ognuna attraverso l'applicazione di un filtro diverso sullo stesso input. Il numero di feature map aumenta generalmente con il crescere della profondità della rete. Questo accade per far fronte alla riduzione dimensionale causata dai livelli di *pooling*. Questi ultimi hanno lo scopo di rendere il modello parzialmente invariante rispetto a piccole traslazioni o disturbi nell'input. Ciò avviene applicando un'operazione di sottocampionamento (*subsampling*) a sottoinsiemi, disgiunti o no, dell'input del livello. Tale operazione può prevedere la selezione del massimo degli elementi del sottoinsieme, oppure della loro media aritmetica. L'operazione di pooling è illustrata in Figura 2.6.

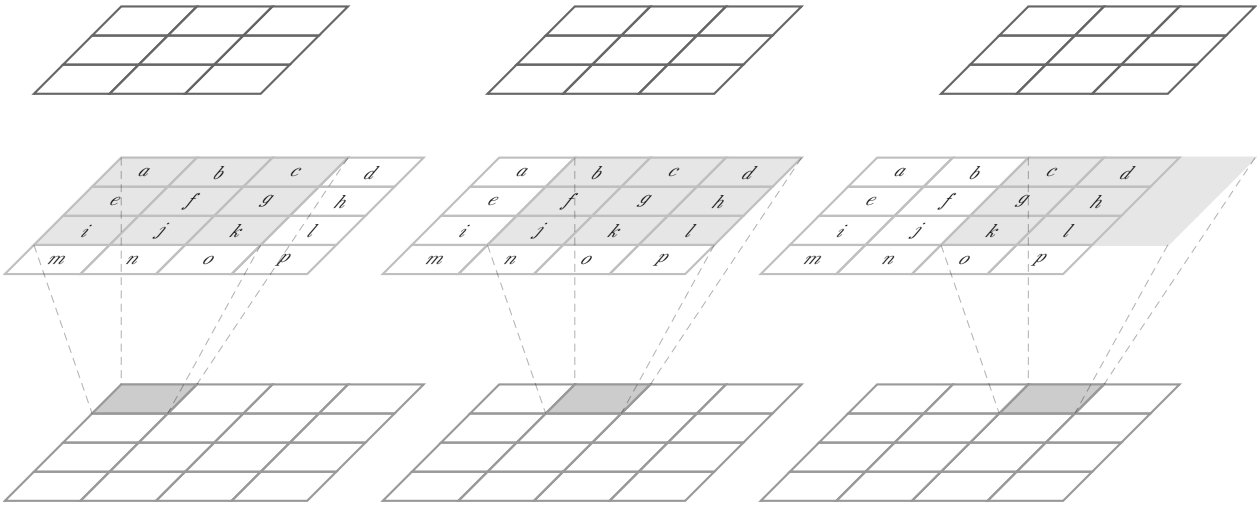


Figura 2.5: Spostamento di un filtro 3×3 lungo una matrice di dimensioni 4×4. Nel primo e secondo riquadro il filtro rientra completamente nelle dimensioni della matrice, mentre nel terzo caso una colonna risulta al di fuori. Gli elementi del filtro che vengono proiettati esternamente alla matrice vengono moltiplicati con valori nulli e non contribuiscono al valore finale della convoluzione.

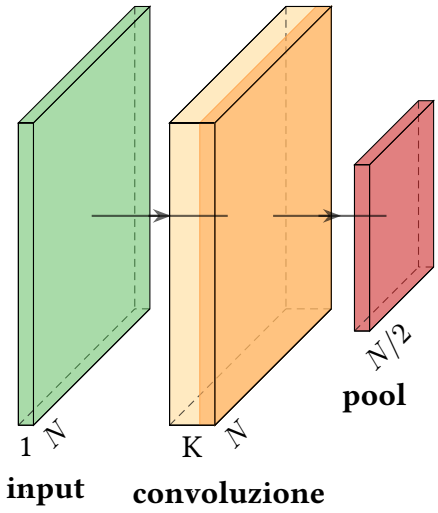


Figura 2.6: Schematizzazione di un semplice blocco convoluzionale a cui è applicata l'operazione di pooling. L'input è bidimensionale e con un solo canale (per esempio l'itensità dei pixel di un'immagine in bianco e nero) e vengono prodotte K feature map. Il livello di pooling dimezza le dimensioni dell'input.

2.6 Regolarizzazione

2.6.1 Overfitting e Underfitting

2.6.2 Regolarizzazione L2

2.6.3 Dropout

Capitolo 3

Architettura Software

In questo capitolo è descritta nel dettaglio l'architettura software sviluppata per il progetto di locallizazione indoor, inclusa la rete neurale, le librerie utilizzate, gli ambienti di sviluppo e gli strumenti che hanno coadiuvato il testing e la sperimentazione del prototipo realizzato.

3.1 TensorFlow

TensorFlow è una libreria open source Python che fornisce metodi e costrutti per definire grafi computazionali e calcolarne automaticamente la derivata. Ciò consente di costruire reti neurali arbitrariamente complesse ed addestrarle tramite le implementazioni degli algoritmi di apprendimento fornite dalla libreria. La versione della libreria utilizzata è TensorFlow 2.2.1, la quale consente di definire funzionalmente modelli di machine learning, interfacciandosi con Keras, una API ad alto livello progettata per essere comprensibile e di facile utilizzo.

3.2 Google Colab

L'ambiente di sviluppo principalmente utilizzato durante il processo di prototipazione è stato Google Colab: una piattaforma cloud hosted progettata da Google per l'esecuzione di Notebook Python e pensata per lo sviluppo di modelli di machine learning. Google Colab

mette a disposizione gratuitamente sistemi di elaborazione con acceleratori computazionali quali GPU e TPU.

3.3 Weights & Biases

Per la gestione sistematica dei test, dei risultati sperimentali e per l'ottimizzazione degli iperparametri del modello, è stato utilizzato Weights & Biases. Esso fornisce una suite di strumenti per il tracking degli esperimenti di machine learning grazie alla registrazione automatica delle metriche di errore durante la fase di addestramento, del salvataggio dei modelli e ai vari tool grafici accessibili dalla piattaforma web di Weights & Biases.

3.4 Architettura della Rete Neurale

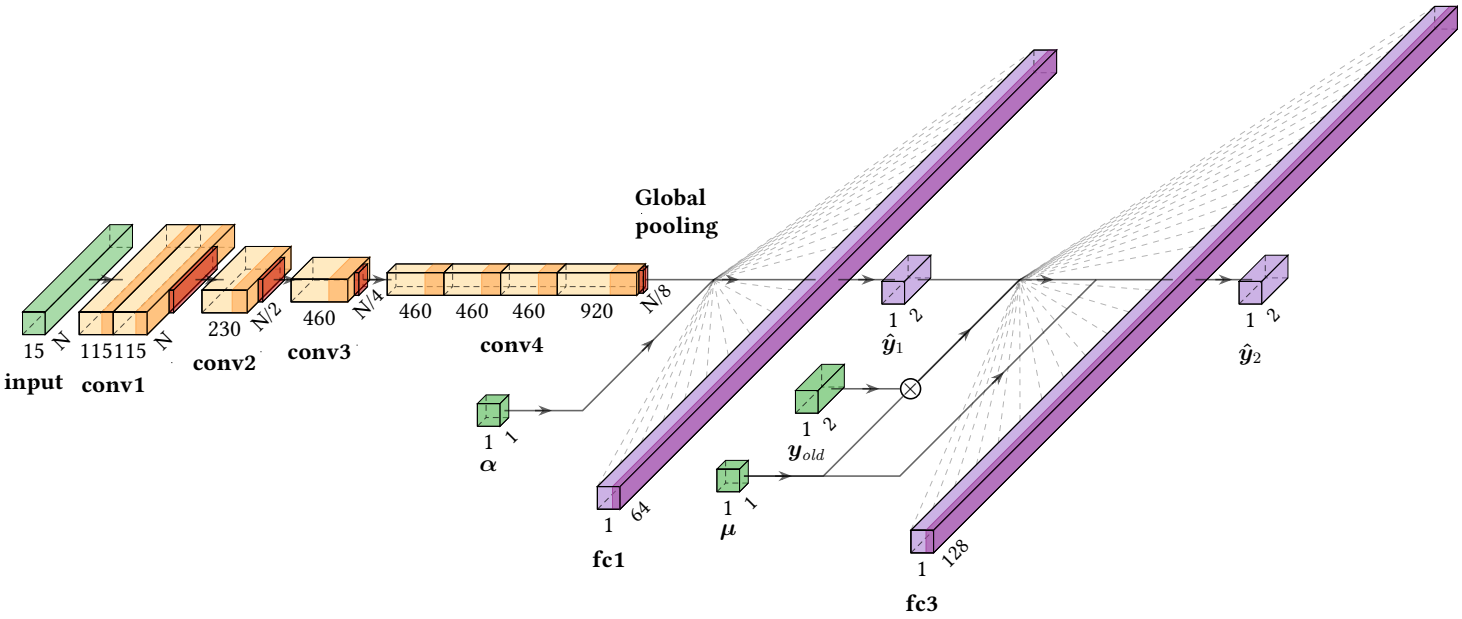


Figura 3.1: Architettura della rete neurale

La rete neurale sviluppata per il problema di localizzazione indoor è illustrata schematicamente in Figura 3.1. Essa consiste in una serie di blocchi convoluzionali seguiti da alcuni livelli di neuroni completamente connessi. Il modello sfrutta, oltre ai segnali RSSI emessi dai beacon, anche due input ausiliari che non sono processati dalla sezione convoluzionale della rete.

3.4.1 Input del Modello

L'input principale del modello è composto da una serie temporale di valori RSSI relativi ai segnali emessi da 15 beacon disposti lungo il perimetro dell'edificio nel quale si sono svolte le sperimentazioni del prototipo. La dimensione temporale dell'input può essere arbitrariamente lunga, poichè una sua variazione determina solamente una differente dimensione dell'asse temporale dell'output della CNN. La sezione convoluzionale della rete si conclude infatti con un livello di pooling globale che consiste nell'estrazione, per ogni feature map prodotta, della media aritmetica dei valori di input lungo la dimensione del tempo. Come si vede in Figura 3.1, infatti, la dimensione dell'output di tale livello è sempre costante e risulta essere 920×1.

A completare l'input del modello sono il valore emesso dal sensore magnetico dello smartphone e l'ultima posizione nota dell'utente all'interno dell'edificio, indicati rispettivamente con α e $\boldsymbol{y}_{old}$. Il valore della bussola è utile per determinare l'orientamento della persona nello spazio, rendendo la rete neurale capace di considerare le variazioni dei segnali BLE dovuti all'assorbimento da parte del corpo dell'utilizzatore dello smartphone. Il secondo input ausiliario è invece utilizzato per correggere eventuali scostamenti rilevanti dell'output della CNN rispetto alla posizione precedente dell'utente. Ci si aspetterebbe infatti che tale posizione non variasse di molto in un lasso di tempo breve.

L'ultima posizione nota dell'utente viene pesata da un coefficiente, anche esso input del modello, che in Figura 3.1 è indicato con la lettera greca μ. Tale valore, cui possiamo riferirci con il termine *coefficiente di memoria residua*, è compreso tra zero e uno, e determina il peso che si vuole dare all'ipotesi di continuità della posizione dell'utente nel tempo. $\mu = 0$

indica l'assenza di tale assunzione, con la conseguente massima riduzione della correzione dell'output della CNN da parte dei livelli successivi, mentre $\mu = 1$ associa il massimo peso su tale ipotesi. Il coefficiente di memoria residua è esso stesso input del livello successivo della rete, il quale riceve anche i valori dell'output ausiliario e di $\mu \cdot \boldsymbol{y}_{old}$.

Il valore della bussola è input del primo livello completamente connesso, insieme all'output della CNN.

3.4.2 Blocco Convoluzionale

La prima parte del modello è una semplice rete neurale convoluzionale unidimensionale. Sebbene sia composto da due dimensioni, quella temporale e quella dei beacon, quest'ultima può essere interpretata come l'insieme dei canali della prima, come nel caso dei canali *r*, *g*, *b* di un'immagine a colori. Ciò permette di applicare l'operazione di convoluzione soltanto lungo l'asse temporale. La CNN proposta è composta da otto blocchi convoluzionali consecutivi così strutturati:

- Operazione di convoluzione sull'output del livello precedente
- Funzione di attivazione ReLU sull'output della convoluzione
- Livello di batch normalization

Un esempio di blocco convoluzionale è illustrato in Figura 3.2, mentre in Figura 3.1 sono mostrati anche i livelli di pooling, rappresentati da una superficie rossa apposta accanto i blocchi convoluzionali.

L'output del secondo, terzo e quarto blocco convoluzionale sono sottoposti ciascuno all'applicazione di un particolare metodo di dropout chiamato *dropout gaussiano*. Esso consiste nell'applicare del rumore moltiplicativo, con distribuzione gaussiana di media unitaria, all'output del blocco convoluzionale. Lo scopo è quello di simulare una corruzione casuale dei dati di addestramento del modello, con l'obiettivo di regolarizzare quest'ultimo. Applicare il rumore all'output dei livelli intermedi della rete, piuttosto che al dataset iniziale,

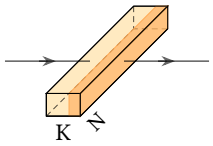


Figura 3.2: Singolo blocco convoluzionale: la parte chiara indica l'operazione di convoluzione, mentre l'ombreggiatura sulla destra illustra la funzione di attivazione ReLU. K è il numero di filtri utilizzati e di conseguenza equivale al numero di feature map prodotte, mentre N è la lunghezza della serie temporale. Il processo di Batch normalization è omesso dalla schematizzazione.

Modello	Loss	MAE	RMSE	MaxAE
Baseline	0.7796	0.3070	0.6716	3.001
No Dropout Gaussiano	0.8911	0.3171	0.7138	3.256

Tabella 3.1: Modello *baseline* messo a confronto con una versione dello stesso che non utilizza il dropout gaussiano. Le metriche fanno riferimento al dataset di *test*.

consente di manipolare più profondamente la rappresentazione dei dati imparata dal modello, rendendolo conseguentemente più robusto rispetto alle variazioni dei segnali di input[4]. Gli effetti dell'utilizzo del dropout gaussiano sono illustrati in Tabella 3.1.

L'output dell'ultimo blocco convoluzionale è infine seguito da un livello di pooling globale e dall'applicazione del dropout.

3.4.3 Uso della Bussola e Output Ausiliario

L'utilizzo dei valori forniti dal sensore magnetico dello smartphone sono giustificati dal voler mitigare il cosiddetto effetto del *body shadowing*. Tale fenomeno si verifica quando un segnale wireless si propaga in un ambiente e collide contro un corpo umano. Tale collisione provoca un decadimento del segnale, il quale arriva disturbato al punto di ricezione. Ciò influisce sulla precisione dei sistemi di localizzazione indoor basati sui valori RSSI dei segnali wireless in modo considerevole, poichè è sufficiente che l'utente volti le spalle a un sottoinsieme dei beacon attivi per introdurre rumore all'interno del sistema. Utilizzando il valore emesso dalla bussola dello smartphone, è possibile rendere il modello consapevole dell'orientamento dell'utente e attenuare il rumore introdotto dal *body shadowing*. I risultati

Modello	Loss	MAE	RMSE	MaxAE
Baseline	0.7796	0.3070	0.6716	3.001
No Bussola	1.619	0.4470	0.9784	4.500

Tabella 3.2: Varie metriche a confronto per i due modelli in esame: *Baseline* è il modello finale, mentre il secondo differisce dal primo soltanto dall'uso dei valori della bussola, che vengono semplicemente scartati. Le metriche si riferiscono al dataset di *test*, ovvero al processo di valutazione successivo alla fase di addestramento.

ottenuti dall'utilizzo di questo input sono illustrati in Tabella 3.2, la quale mette in relazione il modello finale con uno che non considera l'input della bussola.

L'input corrispondente al sensore magnetico è un valore scalare $\alpha \in \mathbb{R}$, con $0 \leq \alpha \leq 360$, in cui il valore 0 indica il nord. Tale dato è fornito, come mostrato in Figura 3.1, al primo livello completamente connesso della rete, insieme all'output della CNN. Questo livello produce un vettore di 64 elementi, il quale diventa input del secondo livello completamente connesso del modello. L'output di quest'ultimo livello, indicato in figura come $\hat{\boldsymbol{y}}_1$, è una coppia di coordinate reali che indica la posizione dell'utente all'interno dell'edificio. Tale previsione è soltanto parziale, in quanto non tiene conto dell'ultima posizione nota dell'utente, ma risulta utile per guidare l'addestramento del modello verso una soluzione meno dipendente da tale input. A questo scopo $\hat{\boldsymbol{y}}_1$ è interpretrato dal modello come output ausiliario.

A ogni output ausiliario è associata una funzione di costo, il cui valore va poi integrato additivamente con quello della funzione di costo dell'output principale. Nel caso della rete neurale in esame si ha:

$$J_1(\boldsymbol{\Theta}) = \frac{1}{m} \sum_{i=1}^{m} \|\hat{\boldsymbol{y}}_{1i} - \boldsymbol{y}_i\|_2^2$$

$$J_2(\boldsymbol{\Theta}) = \frac{1}{m} \sum_{i=1}^{m} \|\hat{\boldsymbol{y}}_{2i} - \boldsymbol{y}_i\|_2^2$$

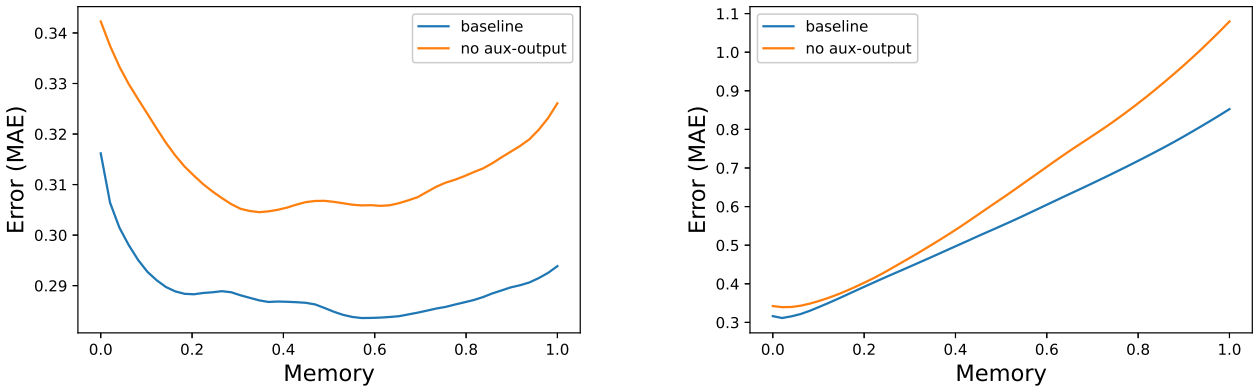
$$J(\boldsymbol{\Theta}) = c_1 J_1(\boldsymbol{\Theta}) + c_2 J_2(\boldsymbol{\Theta})$$

in cui la scelta dei parametri c_1 e c_2 determina il peso di ciascuna funzione di costo nel bilan-

cio dell'errore totale: il modello implementato assegna ai coefficienti i valori $c_1 = \frac{1}{2}, c_2 = 1$. Si noti che tali valori sono completamente arbitrari e rappresentano due iperparametri del modello.

Se l'output $\hat{\boldsymbol{y}}_1$ non pesasse direttamente nella funzione di costo del modello (cioè se fosse $c_1 = 0$), l'output della rete sarebbe troppo condizionato dal valore dell'input ausiliario $\boldsymbol{y}_{old}$. Poichè in fase di raccolta dei dati, la posizione precedente dell'utente non è nota, si assume che tale variabile aleatoria sia distribuita secondo una distribuzione normale centrata nella posizione del campionamento e con deviazione standard σ pari a una piccola costante, indicativa della variabilità del moto di una persona mentre cammina (per esempio $\sigma = 1$). Ciò implica che, qualora fosse $\sum_i \|\boldsymbol{y}_{old\,i} - \boldsymbol{y}_i\|_2^2 < \sum_i \|\hat{\boldsymbol{y}}_{2i} - \boldsymbol{y}_i\|_2^2$, cioè se la distanza media tra l'ultima posizione nota e l'effettiva posizione dell'utente durante il campionamento dei segnali, fosse minore della precisione media ottenibile dal modello, i parametri della rete convergerebbero verso dei valori che tenderebbero a ignorare l'input dei beacon. L'utilizzo della sola posizione precedente dell'utente per esprimere l'output del modello, garantirebbe infatti un errore inferiore rispetto al considerare anche i valori RSSI dei segnali.

Utilizzando l'output ausiliario descritto, pesato con un coefficiente non nullo, viene garantito che lo scenario appena descritto non si verifichi, a patto che il coefficiente selezionato sia sufficientemente grande. In Figura 3.3 il modello descritto è messo a confronto con uno in cui l'output ausiliario non ha peso all'interno della funzione di costo ($c_1 = 0$).



(a) Deviazione standard $\sigma = 1$

(b) Deviazione standard $\sigma = 5$

Figura 3.3: Analisi dell'utilizzo dell'output ausiliario: i grafici mostrano l'andamento dell'errore (MAE) commesso dai due modelli sul dataset di test al variare del coefficiente di memoria residua. La curva blu rappresenta il modello finale, mentre la curva arancione indica il modello senza output ausiliario (cioè in cui $c_1 = 0$). Nel grafico a sinistra, l'input della posizione precedente ($\boldsymbol{y}_{old}$) è perturbato con del rumore gaussiano la cui deviazione standard è pari a $\sigma = 1$. A destra invece $\sigma = 5$. L'input perturbato determina una stima meno precisa dell'ultima posizione nota dell'utente e, con il crescere del coefficiente di memoria residua, entrambi i modelli tendono a sovrastimare l'importanza di tale input. Tuttavia, come si evince dai grafici, l'errore commesso dal modello *baseline* cresce più lentamente ed è sempre minore di quello commesso dal modello senza output ausiliario.

3.4.4 Output del Modello

L'output del modello, indicato in Figura 3.1 come $\hat{\boldsymbol{y}}_2$, consiste in una coppia di coordinate reali che indica la posizione prevista dell'utente all'interno dell'edificio, dopo essere stata opportunamente corretta in base alle conoscenze relative alla precedente locazione dell'utilizzatore.

3.5 Dataset Augmentation e Preprocessing

Poichè il processo di raccolta dei dati è particolarmente dispendioso, sono state impiegate tecniche di *data augmentation* al fine di arricchire il dataset di addestramento. Un insieme di dati più grande corrisponde spesso a una maggiore precisione del modello e consente di

Modello	Loss	MAE	RMSE	MaxAE
No Augmentation	1.2090	0.4072	0.7947	3.481
Replicazione Dataset	1.076	0.3168	0.8344	3.958
Baseline	0.7796	0.3070	0.6716	3.001

Tabella 3.3: Metriche di errore ottenute dal dataset di *test* per tre modelli diversi: il primo non processa in alcun modo i dati originali, il secondo replica il dataset più volte e l'ultimo è il modello finale comprensivo di data augmentation. Come si può notare, l'ultimo modello è il migliore per tutte le metriche esaminate, ma è molto vicino al secondo per quanto riguarda il MAE. Tuttavia la differenza è notevole in termini di RMSE, indicando una varianza più alta nel modello privo di data augmentation.

utilizzare architetture più profonde. Il problema dell'overfitting descresce infatti di intensità con il tendere della cardinalità del dataset di addestramento a infinito.

Le tecniche di data augmentation possono essere interpretate come dei metodi per regolarizzare il modello: esse si basano infatti sull'idea di introdurre rumore all'interno dei dati e rendere quindi la rete neurale più robusta alla corruzione delle informazioni di input. Modellare in qualche modo questo tipo di rumore casuale all'interno del modello permette di limitare la dipendenza dell'output da eventuali variabili latenti non osservate.

In Tabella 3.3 sono mostrati i risultati sperimentali ottenuti dall'utilizzo delle tecniche di data augmentation descritte nel seguito.

3.5.1 Jittering

Si può assumere che l'errore intrinseco della raccolta dei dati, dovuto a limiti tecnologici o a fenomeni non osservati, segua una distribuzione gaussiana a media nulla. Risulta quindi possibile alterare l'input del modello sommandolo con un cosiddetto rumore bianco, ottenendo così un dataset di dimensione doppia rispetto all'originale.

Sia $D = \{\boldsymbol{x}_1, \ldots, \boldsymbol{x}_m\}$ il dataset di addestramento, con $\boldsymbol{x}_i \in \mathbb{R}^n$. Siano poi $\boldsymbol{z}_1, \ldots, \boldsymbol{z}_m$ variabili aleatorie con distribuzione normale multidimensionale tali che:

$$\boldsymbol{z}_i \in \mathbb{R}^n,$$
$$\boldsymbol{z}_i \sim \mathcal{N}(0, \Sigma).$$

La matrice Σ, detta di covarianza, determina la varianza della distribuzione lungo ciascuna dimensione dello spazio. In questo contesto si assume $\Sigma = \sigma^2 I$. Applicando il rumore gaussiano ad ogni input iniziale si ottiene $D' = \{\boldsymbol{x}_1 + \boldsymbol{z}_1, \ldots, \boldsymbol{x}_m + \boldsymbol{z}_m\}$. Il dataset finale consiste nell'unione dei due insiemi $D \cup D'$.

Un esempio illustrativo del jittering è mostrato in Figura 3.4.

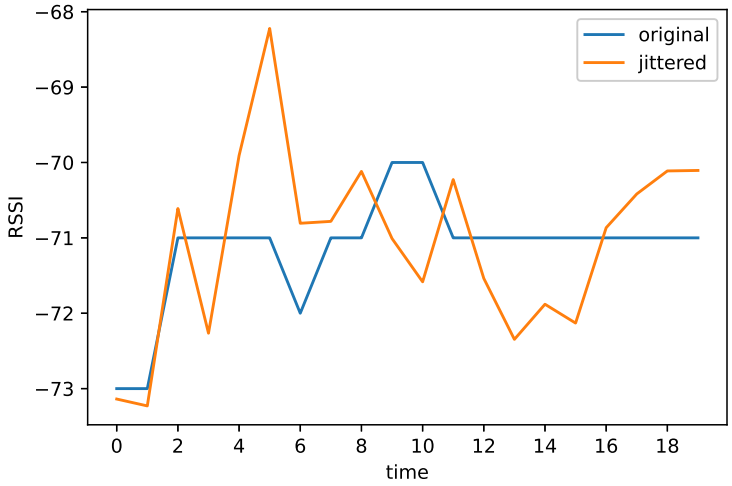


Figura 3.4: Jittering applicato ai segnali emessi da un beacon in un intervallo di tempo limitato. L'asse x descrive il tempo, mentre l'asse y indica la potenza del segnale ricevuto, in dB, al momento t.

3.5.2 Ridimensionamento (Scaling)

La tecnica del ridimensionamento consiste nell'applicare del rumore moltiplicativo costante al segnale di ingresso, mantenendone quindi intatta la struttura, ma modificandone l'ampiezza. L'utilizzo dello *scaling* è giustificato dall'assunzione che, per lo scopo della localizzazione indoor, l'informazione contenuta in una serie temporale di segnali è invariante rispetto a piccole variazioni dell'ampiezza dei segnali, poichè queste potrebbero essere causate da fenomeni esterni non osservati.

Utilizzando la stessa notazione della sezione precedente, definiamo una variabile aleatoria $c \sim \mathcal{N}(1, \sigma^2)$ con distribuzione gaussiana a media unitaria e varianza σ^2. Il singolo

elemento del dataset a cui è applicato il ridimensionamento è quindi definito come:

$$\boldsymbol{x}' = c\boldsymbol{x}$$

Un esempio di scaling applicato ai segnali emessi da un beacon è illustrato in Figura 3.5.

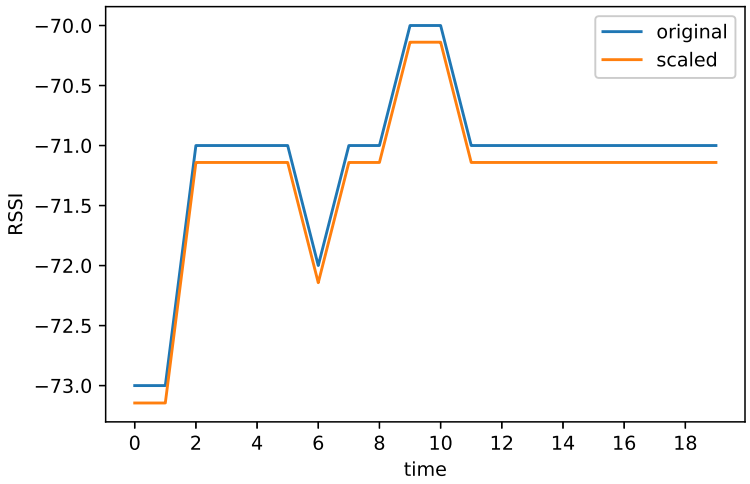


Figura 3.5: Illustrazione grafica del ridimensionamento di un segnale.

3.5.3 Magnitude Warping

Data la natura delle perturbazioni a cui sono soggetti i segnali dei beacon nel contesto della localizzazione indoor, è lecito pensare che tali distorsioni possano verificarsi anche in periodi particolarmente limitati del tempo e con una intensità relativamente elevata. Tale assunzione permette di sviluppare una tecnica di data augmentation a cui possiamo dare il nome di *magnitude warping*, la quale prevede di deformare l'intensità del segnale in punti limitati della serie temporale, applicando del rumore moltiplicativo a una porzione di sottosequenze, non necessariamente disgiunte, del segnale di input. In Figura 3.6 è mostrato graficamente l'utilizzo di tale tecnica. Il Listato 3.1 descrive invece i dettagli implementativi del magnitude warping.

Algorithm 3.1 Descrizione algoritmica del funzionamento del magnitude warping

function MAGNITUDEWARPING($\boldsymbol{v}$, σ^2, $MaxLength$, $MaxPeaks$)
  $\boldsymbol{v}' \leftarrow \boldsymbol{v}$
  $sample\ n \sim Uniform(1,\ MaxPeaks)$
  for $peak = 0, \ldots, n$ **do**
    $sample\ \ell \sim Uniform(1,\ MaxLength)$
    $sample\ p \sim Uniform(0,\ length(\boldsymbol{v}) - \ell)$
    $sample\ m \sim \mathcal{N}(1, \sigma^2)$
    for $i = p, \ldots, p + \ell$ **do**
      $\boldsymbol{v}'[i] \leftarrow m \cdot \boldsymbol{v}'[i]$
  return $\boldsymbol{v}'$

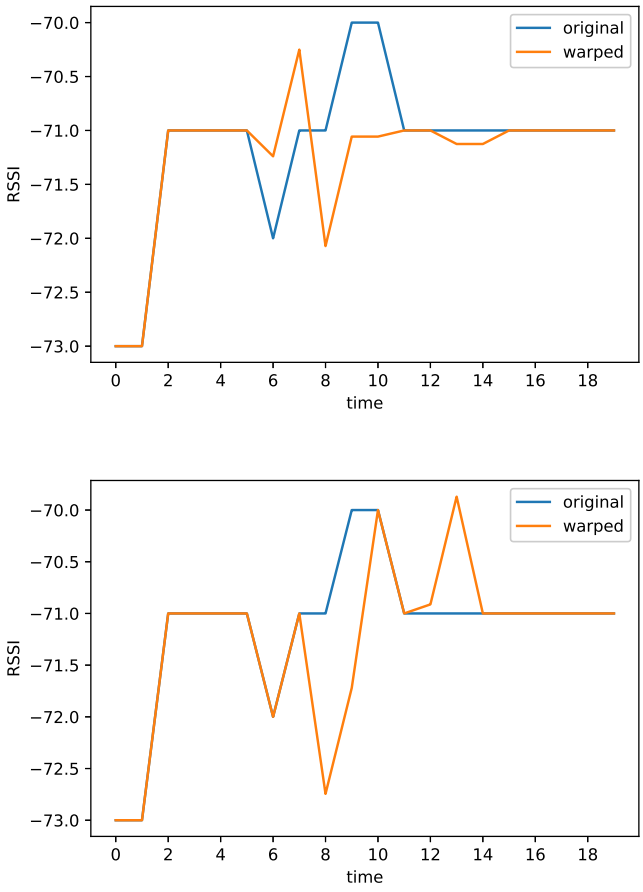


Figura 3.6: Esempi grafici di magnitude warping applicato a un segnale.

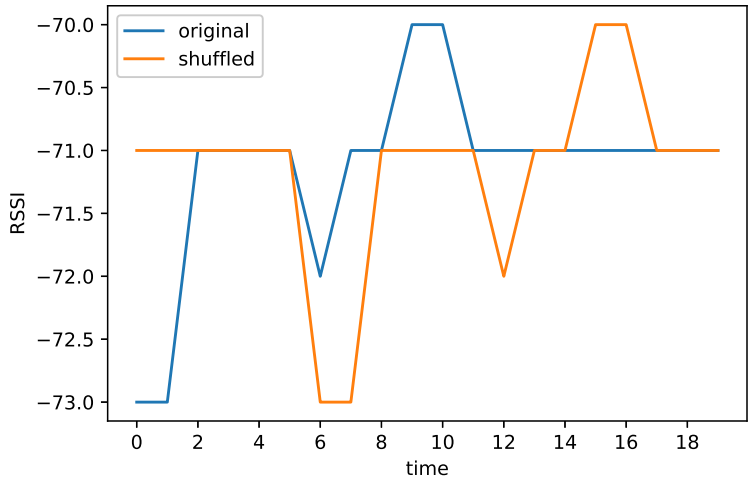


Figura 3.7: Shuffling di sottoinsiemi applicato a una serie temporale.

3.5.4 Permutazione di Sottoinsiemi (Subset Shuffling)

Un'altra assunzione che è possibile fare riguardo l'input del problema è che l'ordine con cui si verificano variazioni all'interno della serie temporale non dovrebbe avere peso nell'estrazione di informazioni o di pattern dai segnali. Per questo motivo è stata implementata una tecnica di *shuffling* in cui sottoinsiemi dell'input vengono scambiati tra di loro secondo permutazioni casuali. Tale metodo è illustrato in Figura 3.7

3.5.5 Deattivazione Selettiva

Per rendere il modello più robusto alla eventuale perdita di segnali di alcuni beacon, è stata introdotta una tecnica di data augmentation che prevede la conversione dei valori della serie temporale di un beacon in un *placeholder* che indica la mancata ricezione di segnale. Nell'ambito di questo progetto, tale valore è -200.

Questa strategia è simile a quella del dropout, ma al posto del valore nullo, l'input viene rimpiazzato con il placeholder scelto. A gestire la frequenza con cui un beacon viene disattivato all'interno del dataset è un iperparametro di tipo percentuale.

3.6 Addestramento del Modello

Il modello descritto è stato addestrato su un dataset di 279457 elementi, ottenuti applicando le tecniche di data augmentation introdotte precedentemente. L'algoritmo utilizzato per l'addestramento è una variante del metodo del gradiente stocastico con momento, che prende il nome di *Adam*. Questo ottimizzatore gestisce autonomamente l'iperparametro della rete che determina l'entità dell'aggiornamento dei parametri del modello: il cosiddetto *learning rate*. A coadiuvare l'algoritmo è stato inserito un metodo, guidato dal dataset di validazione, che riduce esponenzialmente il learning rate ogni volta che la procedura di addestramento raggiunge uno stallo.

Un ulteriore metodo di regolarizzazione introdotto è quello dell'*Early Stopping*, il quale termina precocemente l'addestramento quando una determinata metrica sul dataset di validazione non migliora per un periodo di tempo predefinito. Quando ciò accade, la configurazione del modello che ha ottenuto i migliori risultati viene ripristinata. Empiricamente si è visto che il metodo dell'early stopping riduce l'errore commesso dal modello sul dataset di validazione, ma è stato anche mostrato come tale algoritmo agisca da regolarizzatore per la rete in un modo simile a quello della regolarizzazione L2[1].

Infine, al termine della procedura di addestramento, il modello ottenuto viene valutato sul dataset di test e i risultati vengono salvati per essere in futuro analizzati e confrontati con altri modelli.

3.7 Ensembling

Al fine di ottenere un errore di generalizzazione minore e migliorare ulteriormente l'affidabilità del sistema di localizzazione indoor, sono stati addestrati indipendentemente, e con differenti configurazioni di iperperparamentri, vari modelli, i quali sono stati successivamente aggregati tramite il metodo dell'*ensembling*. Tale tecnica consiste nell'unire le previsioni di modelli diversi per formarne una il cui errore sia statisticamente più basso, a patto che gli errori commessi dai singoli modelli siano sufficientemente non correlati[2]. In

Modello	MAE	RMSE	MaxAE
Baseline	0.3070	0.6716	3.001
Ensemble	0.2592	0.5536	2.4693

Tabella 3.4: Confronto tra il modello *baseline* e un ensemble di cinque modelli addestrati individualmente. Le metriche di errore si riferiscono al dataset di test.

un problema di regressione, ad esempio, per aggregare l'output di un insieme di modelli è possibile utilizzare la media aritmetica delle varie previsioni.

Un ulteriore vantaggio dell'utilizzo di un ensemble di modelli, è la possibilità di ottenere un'indicazione sulla variabilità della previsione, calcolando la deviazione standard dell'output dei modelli. Ciò fornisce un'informazione sulla confidenza dell'ensemble relativamente alla previsione di un determinato output. Tale informazione può essere poi sfruttata al fine di stabilizzare il modello con un filtro di Kalman, come descritto nel Paragrafo 4.4.2. Utilizzando un ensemble di cinque modelli è stato possibile ottenere i risultati illustrati in Tabella 3.4.

Capitolo 4

Applicazione Mobile

4.1 Flutter

4.2 Planimetrie e Poligoni

4.3 Backend TensorFlow

4.3.1 TensorFlow Lite

4.3.2 Implementazione del Bridge di Comunicazione

4.4 Stabilizzazione del Modello

4.4.1 Utilizzo di Sensori Inerziali

4.4.2 Filtro di Kalman

Capitolo 5

Conclusioni

5.1 Risultati Sperimentali

5.1.1 Metriche di Errore: MSE, MAE, MaxAE

5.2 Lavori futuri

5.2.1 Input a Lunghezza Variabile

5.2.2 Reti Neurali Residuali

5.2.3 Variational Autoencoder: Generazione di nuovi dati

5.2.4 Transfer Learning

5.2.5 Input Masking e Ricostruzione dei Segnali

5.2.6 Transformers per Problemi di Regressione

5.2.7 Simulatore BLE

5.2.8 Posizionamento Magnetico

Bibliografia

[1] Christopher Bishop. "Regularization and Complexity Control in Feed-forward Networks". In: *Proceedings International Conference on Artificial Neural Networks ICANN'95*. Vol. 1. EC2 et Cie, gen. 1995, pp. 141–148. URL: https://www.microsoft.com/en-us/research/publication/regularization-and-complexity-control-in-feed-forward-networks/.

[2] Ian Goodfellow, Yoshua Bengio e Aaron Courville. *Deep Learning*. http://www.deeplearningbook.org. MIT Press, 2016.

[3] Allan Pinkus Moshe Leshno Vladimir Ya. Lin e Shimon Schocken. "Multilayer Feedforward Networks With a Nonpolynomial Activation Function Can Approximate Any Function". In: *Neural Networks* (1993).

[4] Ben Poole, Jascha Sohl-Dickstein e Surya Ganguli. "Analyzing noise in autoencoders and deep networks". In: *CoRR* abs/1406.1831 (2014). arXiv: 1406.1831. URL: http://arxiv.org/abs/1406.1831.