Exploiting the Use of Convolutional Neural Networks for Localization in Indoor Environments

Bruno V. Ferreira[a,e], Eduardo Carvalho[b,e], Mylena R. Ferreira[a,e], Patrícia A. Vargas[c], Jó Ueyama[d], Gustavo Pessin[e]

[a] *Institute of Technology, Federal University of Pará, Belém, PA, Brazil*
[b] *SENAI Institute of Innovation in Minerals Technologies, Belém, PA, Brazil*
[c] *Robotics Laboratory, School of Mathematical and Computer Sciences (MACS) Heriot-Watt University, Edinburgh, UK*
[d] *Institute of Mathematics and Computer Science (ICMC) University of São Paulo (USP), São Carlos, São Paulo, Brazil*
[e] *Applied Computing Lab, Vale Institute of Technology, Belém, PA, Brazil*
{bruno.ferreira, mylena.ferreira}@pq.itv.org, eduardo.isi@sesipa.org.br, p.a.vargas@hw.ac.uk, joueyama@icmc.usp.br, gustavo.pessin@itv.org

---

Abstract

Indoor localization has been an active research area for the last two decades. A great number of sensors have been applied in the task of localization, some with high computational and energy demands, like laser beams, or with issues related to the coverage area, for example, by making use of images obtained by a network of cameras. A different approach, that presents less energy demands and a wide area of coverage, can be created by means of signal strength of wireless networks. The open issue with signal strength is it high instability due to interferences, attenuation and fading; which, in general, makes the localization systems to present less than desired accuracy. In this article we exploit the use of Convolutional Neural Networks (ConvNets) in the task of localization. The main motivation behind the employment of ConvNets is its inherent ability of feature extraction, which we believe can deal better with the noise without a filtering step. We evaluate how ConvNets can be employed and which are the best topologies that lead to the lowest errors.

Keywords: Indoor Localization, Convolutional Neural Networks, Wi-Fi

---

1. Introduction

Localization in indoor environments, such as buildings or underground mines, has been focus of a great number of researches in the last few years. Localization can be useful in a several number of tasks, like for improving rescue in case of fires in buildings or landslides in caves or underground mines. It can also be a source of information in shopping malls to understand people behavior or in hospitals to track patients with mental disorders. Although localization is useful information, the techniques employed to measure it in indoor environments still presents a number of open questions, mainly related to accuracy and usability.

According to Harris and colleagues (Harris et al., 2014), in the period of 2006-2010, there were more than 2,500 deaths in underground mines, being 75 in the US alone. Moreover, according to the US Fire Administration (US Fire Administration, 2014), in the period 2003-2012, there were more than 3.5 million of fires in private residences and near 1 million occurrences in non-residential buildings, bringing thousands of fatalities and injured people. An easy-to-use localization system could help the rescue of the impacted people and also could be employed by the rescuers to improve their own safety. Miners and speleologists are other two groups of people in which the system could help improving the safety in their regular activities.

A wide range of technologies have been evaluated for the localization of people in outdoor environments in the last few decades, such as (i) global positioning systems (GPS), (ii) inertial measurement units (IMU), (iii) vision sensors, or a combination of these. However, localization in indoor environments is still an open question due to the complexity of indoor environments (Elnahrawy et al., 2004; Ladd et al., 2004). On the other hand, robot localization techniques have achieved good results in indoor environments, even though the use of these technologies is not feasible for the localization of people on account of the high computational cost of processing the data gathered by the sensors that are used. Thus, a different type of methodology have been adopted that make use of signal strength of wireless devices, driven by the large number of devices that make use of it for communication (Yoo et al., 2014; BÃćce and Pignolet, 2015; Pessin et al., 2014).

A current issue while dealing with signal strength is its vulnerability to interferences, fading and attenuation. Those characteristics add noise to the signal, making it somehow difficult to be employed for high accuracy localization systems. Machine learning techniques, like multilayer artificial neural networks and support vector machines, among other, have been employed to estimate localization due to its inherent learning and generalization capabilities. It is expected that the learning and generalization capabilities allows the estimation of the localization to have a good accuracy despite the noise. Although, as can be seen in the work by Carvalho and colleagues (Carvalho et al., 2016), the use of a filtering step (moving average) before the use of the machine learning techniques improves the accuracy of the system, however, it adds more complexity to the system.

In this article we exploit the use of Convolutional Neural Networks (ConvNets) in the task of localization. The main motivation behind the employment of ConvNets is its inherent ability of feature extraction, which we believe can deal better with the noise without a filtering step. We evaluate how ConvNets can be employed and which are the best topologies that lead to the lowest errors. The article is organized into the following sections: In Section 2 we describe concepts about Convolutional Neural Networks. In Sections 3 we present the environment employed for indoor localization and the proposed ConvNet model that deals with a time series of signal strength from access points. The results are presented and discussed in Section 4. We finish the paper presenting the conclusions and directions for future work.

2. Deep Neural Networks

Before the development of the deep learning field, it was common that big neural network use to suffer of training inefficiency due to the problem of the vanishing gradient (Dalto, 2015). Two are the more common neural networks employed within the deep learning concepts: (i) Convolutional Neural Networks (ConvNets) and (ii) Recurrent Neural Networks (RNNs).

Convolutional Neural Networks, as described by Lecun, Bottou, Bengio and Haffner (Lecun et al., 1998), are neural networks that leverage the biological concept of receptive fields. It is often divided in two parts: convolutional layers followed by a multilayer perceptron layer. The convolutional layer performs inherent feature extraction and the MLP layer is responsible for the classification or regression. ConvNets are commonly employed in tasks such image and sound classification (Sermanet et al., 2013; Abdel-hamid et al., 2013; Krizhevsky et al., 2012; LeCun et al., 2015). Recurrent Neural Networks are a second type of artificial neural networks employed for deep learning process. It is a class of multi-

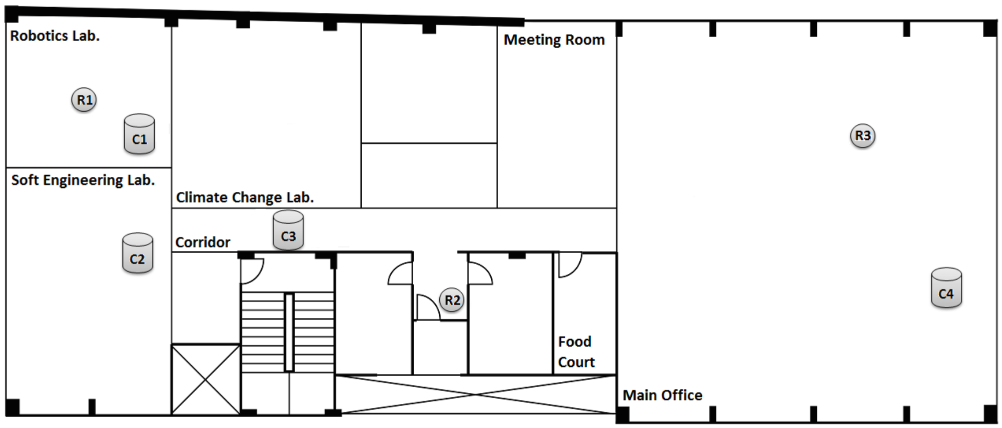

Figure 1: The testing environment: There are 3 routers (R1 to R3) which broadcast RSSI signals and 4 rooms were collections were carried out (C1 to C4). In each room (C1 to C4) nine different collections were performed taking $\approx$ 3 to 4 minutes each (a total of 30 minutes in each room, and 2 h taking into account the four rooms). The RSSI are employed as inputs for the machine learning position estimation.

player perceptron networks in which some neurons have connections with former neurons, as in a directed cycle; it makes the RNNs to exhibit dynamic temporal behavior (Sak et al., 2014). RNNs are commonly employed in handwriting recognition or speech recognition (Pham et al., 2014; Graves and Jaitly, 2014). Due to recurrent connections, it is said that RNNs present temporal relationship among the inputs, being some sort of memory.

In this work we propose and evaluate Convolutional Neural Networks in the task of localization due to its inherent ability of feature extraction. Our proposed ConvNet topology is described in Section 3.

3. Environment and Methods for Indoor Localization

As previously mentioned, in this article we exploit the use of ConvNets to estimate localization in indoor environment by means of signal strength (RSSI)

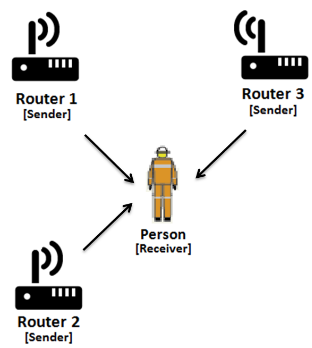

Figure 2: Architecture of the system: a person must have a device that receives RSSI from routers and employs machine learning to estimate its location.

of wireless nodes. Fig. 1 shows the environment: There are 3 routers (R1 to R3) which broadcast RSSI signals and 4 rooms were collections were carried out (C1 to C4). In each room (C1 to C4) nine different collections were performed taking $\approx$ 3 to 4 minutes each (a total of 30 minutes in each room, and 2 h taking into account the four rooms). Each scan of the networks takes approximately 0.4 seconds; hence, in 2 h of collection the dataset for the learning process of the ConvNet has a total of 17617 records from signal strength. The dataset was divided in a proportion of 70% for training and 30% for validation.

Figure 2 shows the general architecture of the system where a person must have a device that receives RSSI from routers and employs machine learning to estimate its location. Figure 3 shows a simplified model of the system, taking into account the inputs and outputs of intelligent system. The ConvNet receives as inputs a time series of signal strength (10 reads) from 3 access points. In this sense, the ConvNet has an input layer of 30 values. The output layer presents 4 logical neurons, representing each room of interest.

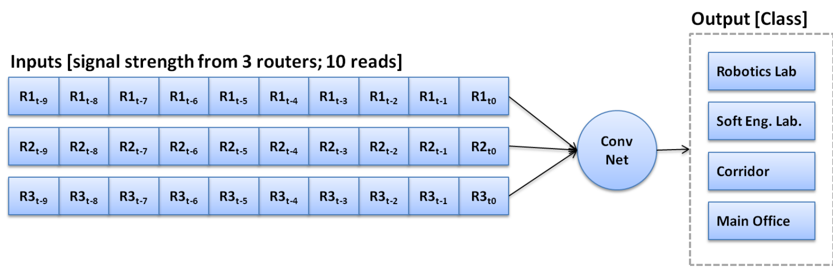

Figure 3: Simplified model of the inputs and outputs of the ConvNet. The ConvNet receives as inputs signal strengths and it is trained to output the class related to each room of interest.

Related to the application of the Convolutional Neural Networks, as previously mentioned, it presents as inputs a time series with 30 values (10 values from 3 routers) and 4 values as outputs (one logical value to each room). Our proposed ConvNet is build to deal with time series, and is based on the work of by Zheng ans colleagues Zheng et al. (2014). The initial architecture is represented in Fig. 4 and it consists of two sets of layers: 1D convolution, 1D Relu and MaxPool; two fully connected layers with Relu and Dropout between them and an output layer with SoftMax. Relu was used because it speeds up the training over classical activation functions Krizhevsky et al. (2012) as sigmoid and hyperbolic tangent was used Dropout to increase the spread of the network and avoid overfitting Srivastava et al. (2014).

The learning rate used was 0.001 and momentum of 0.005 using Nesterov Momentum Sutskever et al. (2013). The amount of trainable parameters ranged from approximately 1700 to 2000. All of the network weights are initialized randomly.

We evaluated 27 different architectures, taking into account {5, 10, 15} filters in the first layer, {5, 10, 15} filters in the second layer and 6x6, 12x12, 24x24

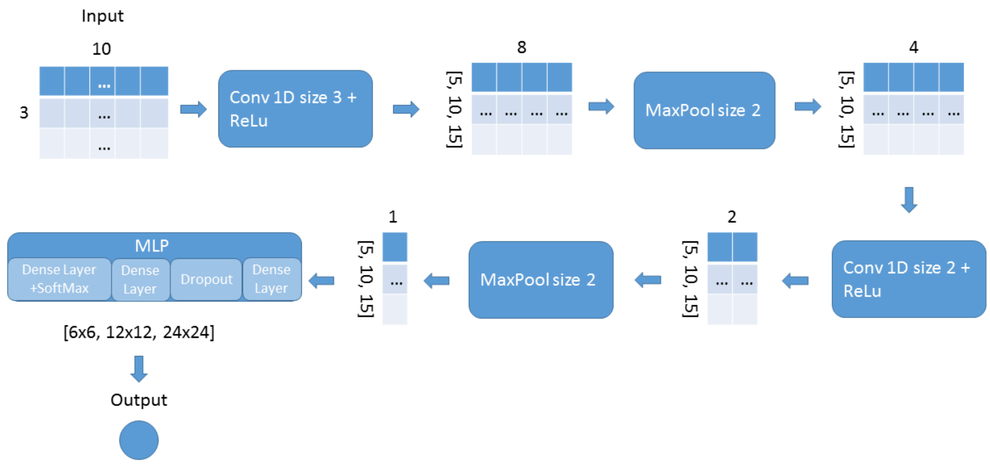

Figure 4: Convolutional Neural Network architecture, displaying the layers and feature maps generated. Between each feature map is show a layer with the size of the filter. For each the feature maps, in brackets is the number of filters in the convolution layer and for the MLP layer is the number of neurons in each layer dense layer.

neurons in the MLP classification layer. Each network was trained with 20,000 epochs and the result aggregates 10 runs of each architecture.

4. Results

Fig. 5 shows the results for the different ConvNet architectures. We can see that architectures with 5 filters in any convolution layer or 6x6 neurons in the hidden layers present a large dispersion in the results. All the best sets were obtained by ConvNets with two convolutional layers with at least 10 filters plus 2 fully connected layers with at least 12x12 neurons each. Taking this into account, the architecture with 10 filters on both convolutional layers and 12x12 neurons in each hidden layer (c10c10d12) was considered the most appropriate for the task of classifying, since it is the smallest architecture that present best classification

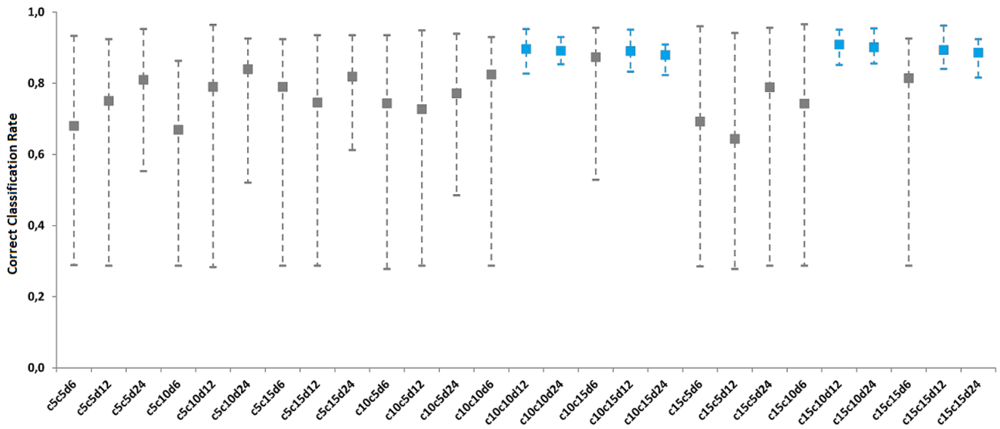

Figure 5: Results of correct classification rates for the evaluated ConvNet topologies. Each line presents the average of ten runs plus the max and min of the set. In blue we show the best results, i.e. the ConvNets that presented correct classification rates with lower dispersion. All the best sets were obtained by ConvNets with two convolutional layers with at least 10 filters plus 2 fully connected layers with at least 12 neurons each.

rates.

Fig. 6 shows the confusion matrix. It is worthwhile to mention that the errors occurred in neighbor regions; by means of the confusion matrix we can see the errors that occurred in regions 1 and 2 (Fig. 1), which represents two neighbor rooms (Robotics Lab. and Soft Eng. Lab.).

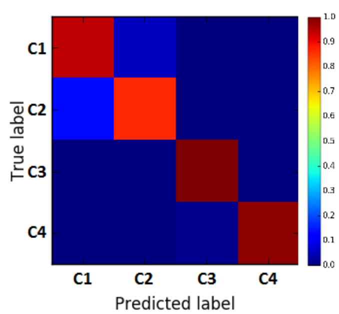

Figure 6: Confusion matrix of the best run of the architecture c10c10d12.

5. Conclusion and Future work

In this paper we exploited the use of Convolutional Neural Networks to perform localization of people in indoors environments. Several architectures were evaluated seeking to understand which topology could obtain best results. As seen in this paper, Convolutional Neural Networks are widely used for image classification, since it presents an inherent feature extraction characteristic. We employ the idea of the inherent feature extraction as a noise-removing filter. Our developments lead to a ConvNet that receives a time series of signal strength values from different access points and perform classification. The different architectures, with several different layer showed that there are topologies more suitable to solve the problem.

Future studies should address other Deep Learning concepts, by exploiting Recurrent Neural Nets and Denoising Autoencoders in the task of localization. Other open question is related to the employed sensors: there is nowadays a plethora of wireless sensors with different frequencies. In this sense, the question of witch sensor (or sensors) could provide the best set to improve the accuracy of the system is a field that deserves attention.

6. Acknowledgments

The authors would like to acknowledge the following colleagues: Geraldo Pereira (USP), Bruno FaiÃğal (USP), Gerson Serejo (ITV) and Helder Arruda (ITV) due to their time discussing ideas. Furthermore, the authors would like to thank Prof. Dr. Cleidson R. B. Souza (ITV) and Dr. Joner Oliveira Alves (SENAI) due to inspirational thoughts and to several aids to the project. Finally, the authors acknowledge the financial support from the “Edital SENAI SESI de InovaÃğÃčo (CNI)”, Vale S. A. and CNPQ by means of the call 59/2013 MCTI/CT-Info/CNPq, process 440880/2013-0.

References

Abdel-hamid, O., Deng, L., Yu, D., 2013. Exploring Convolutional Neural Network Structures and Optimization Techniques for Speech Recognition. In: 14th Annual Conference of the International Speech Communication Association, (INTERSPEECH 2013). No. August. pp. 3366–3370.

BÃćce, M., Pignolet, Y. A., Oct 2015. Lightweight indoor localization system. In: 2015 8th IFIP Wireless and Mobile Networking Conference (WMNC). pp. 160–167.

Carvalho, E., FaiÃğal, B. S., Filho, G. P. R., Vargas, P. A., Ueyama, J., Pessin, G., March 2016. Exploiting the use of machine learning in two different sensor network architectures for indoor localization. In: 2016 IEEE International Conference on Industrial Technology (ICIT). pp. 652–657.

Dalto, M., 2015. Deep neural networks for time series prediction with applications in ultra-short-term wind forecasting. Industrial Technology (ICIT), 2015 IEEE International Conference on, 1657–1663.
URL `http://ieeexplore.ieee.org/xpl/articleDetails.jsp?arnumber=7125335`

Elnahrawy, E., Li, X., Martin, R. P., Oct 2004. The limits of localization using signal strength: a comparative study. In: Sensor and Ad Hoc Communications and Networks, 2004. IEEE SECON 2004. 2004 First Annual IEEE Communications Society Conference on. pp. 406–414.

Graves, A., Jaitly, N., 2014. Towards end-to-end speech recognition with recurrent neural networks. In: ICML. Vol. 14. pp. 1764–1772.

Harris, J., Kirsch, P., Shi, M., Li, J., Gagrani, A., Krishna, E. S., Tabish, A., Arora, D., Kothandaraman, K., Cliff, D., Others, 2014. Comparative analysis of coal fatalities in Australia, South Africa, India, China and USA, 2006-2010. In: 14th Coal Operator's Conference.

Krizhevsky, A., Sutskever, I., Hinton, G. E., 2012. ImageNet Classification with Deep Convolutional Neural Networks. Advances In Neural Information Processing Systems, 1–9.

Ladd, A. M., Bekris, K. E., Rudys, A. P., Wallach, D. S., Kavraki, L. E., June 2004. On the feasibility of using wireless ethernet for indoor localization. IEEE Transactions on Robotics and Automation 20 (3), 555–559.

LeCun, Y., Bengio, Y., Hinton, G., 2015. Deep learning. Nature 521 (7553), 436–444.

Lecun, Y., Bottou, L., Bengio, Y., Haffner, P., Nov 1998. Gradient-based learning applied to document recognition. Proceedings of the IEEE 86 (11), 2278–2324.

Pessin, G., Osório, F. S., Ueyama, J., Wolf, D. F., Moioli, R. C., Vargas, P. A., 2014. Self-localisation in indoor environments combining learning and evolution with wireless networks. In: Proceedings of the 29th Annual ACM Symposium on Applied Computing. ACM, pp. 661–666.

Pham, V., Bluche, T., Kermorvant, C., Louradour, J., 2014. Dropout improves recurrent neural networks for handwriting recognition. In: Frontiers in Handwriting Recognition (ICFHR), 2014 14th International Conference on. IEEE, pp. 285–290.

Sak, H., Senior, A. W., Beaufays, F., 2014. Long short-term memory recurrent neural network architectures for large scale acoustic modeling. In: INTERSPEECH. pp. 338–342.

Sermanet, P., Eigen, D., Zhang, X., Mathieu, M., Fergus, R., LeCun, Y., 2013. OverFeat: Integrated Recognition, Localization and Detection using Convolutional Networks. arXiv preprint arXiv, 1312.6229.
URL `http://arxiv.org/abs/1312.6229`

Srivastava, N., Hinton, G. E., Krizhevsky, A., Sutskever, I., Salakhutdinov, R., 2014. Dropout : A Simple Way to Prevent Neural Networks from Overfitting. Journal of Machine Learning Research (JMLR) 15, 1929–1958.

Sutskever, I., Martens, J., Dahl, G. E., Hinton, G. E., 2013. On the importance of initialization and momentum in deep learning. Jmlr W&Cp 28 (2010), 1139–1147.

URL http://dblp.uni-trier.de/db/conf/icml/icml2013.html{\#}SutskeverMDH13

US Fire Administration, 2014. www.usfa.fema.gov/data/statistics/#tab-3.

Yoo, J., Kim, T., Provencher, C., Fong, T., 2014. WiFi Localization on the International Space Station. In: IEEE Symposium on Intelligent Embedded Systems (IES). IEEE, pp. 21–26.

Zheng, Y., Liu, Q., Chen, E., Ge, Y., Zhao, J. L., 2014. Time series classification using multi-channels deep convolutional neural networks. In: Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics). Vol. 8485 LNCS. pp. 298–310.