



Applied Artificial Intelligence

An International Journal

ISSN: 0883-9514 (Print) 1087-6545 (Online) Journal homepage: <https://www.tandfonline.com/loi/uaai20>

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

Bruno V. Ferreira, Eduardo Carvalho, Mylena R. Ferreira, Patrícia A. Vargas, Jó Ueyama & Gustavo Pessin

To cite this article: Bruno V. Ferreira, Eduardo Carvalho, Mylena R. Ferreira, Patrícia A. Vargas, Jó Ueyama & Gustavo Pessin (2017) Exploiting the Use of Convolutional Neural Networks for Localization in Indoor Environments, Applied Artificial Intelligence, 31:3, 279-287, DOI: [10.1080/08839514.2017.1316592](https://doi.org/10.1080/08839514.2017.1316592)

To link to this article: <https://doi.org/10.1080/08839514.2017.1316592>



Published online: 12 May 2017.



Submit your article to this journal [↗](#)



Article views: 434



View related articles [↗](#)



View Crossmark data [↗](#)



Citing articles: 1 View citing articles [↗](#)



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

Bruno V. Ferreira^{a,b,c}, Eduardo Carvalho^{b,c}, Mylena R. Ferreira^{a,b,c},
Patrícia A. Vargás^d, Jó Ueyama^e, and Gustavo Pessin^b


^aInstitute of Technology, Federal University of Pará, Belém, PA, Brazil; ^bApplied Computing Laboratory, Instituto Tecnológico Vale, Belém, PA, Brazil; ^cIoT Lab, SENAI Institute of Innovation in Minerals Technologies, Belém, PA, Brazil; ^dRobotics Laboratory, School of Mathematical and Computer Sciences (MACS), Heriot-Watt University, Edinburgh, UK; ^eInstitute of Mathematics and Computer Science (ICMC), University of São Paulo (USP), São Carlos, São Paulo, Brazil

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 (e.g. 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, which presents less energy demands and a wide area of coverage, can be created by means of the signal strength of wireless networks. The open issue with signal strength is its 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 identify the best topologies that lead to the lowest errors.

Introduction

Localization in indoor environments, such as buildings or underground mines, has been the focus of a great number of research works in the last few years. Localization can be useful in several 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 provides useful information, the techniques employed to measure it in indoor environments still present a number of open questions mainly related to accuracy and usability.

CONTACT Gustavo Pessin ✉ gustavo.pessin@itv.org  Applied Computing Laboratory, Instituto Tecnológico Vale, R. Boaventura da Silva, 955 Belém, PA 66055-090, Brazil.

Color versions of one or more of the figures in the article can be found online at <http://www.tandfonline.com/UAAI>.

© 2017 Taylor & Francis

According to Harris and colleagues (Harris et al. 2014), in the period of 2006-2010, there were more than 2500 deaths in underground mines, with 75 being in the United States 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 fire occurrences in private residences and nearly 1 million 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 the 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, Li, and Martin 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 has been adopted that makes use of the signal strength of wireless devices and is driven by the large number of devices that make use of it for communication (Bâce and Pignolet 2015; Pessin et al. 2014; Yoo et al. 2014).

A current issue while dealing with signal strength is its vulnerability to interferences, fading and attenuation. These 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 will allow 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, 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 would lead to the lowest errors. The article is organized into the following sections: In “Deep neural networks” section, we describe concepts about ConvNets. In “Environment and methods for indoor localization” section,

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 “Results” section. We finish the paper by presenting the conclusions and directions for future work.

Deep neural networks

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

ConvNets, 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 into two parts—convolutional layers followed by a multilayer perceptron (MLP) layer. The convolutional layer performs inherent feature extraction, while the MLP layer is responsible for the classification or regression. ConvNets are commonly employed in tasks such as image and sound classification (Abdel-Hamid, Deng, and Yu 2013; Krizhevsky, Sutskever, and Hinton 2012; LeCun, Bengio, and Hinton 2015; Sermanet et al. 2013). RNNs are the second type of artificial neural networks employed for the deep learning process. It belongs to a class of multilayer 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, Senior, and Beaufays 2014). RNNs are commonly employed in handwriting recognition or speech recognition (Graves and Jaitly 2014; Pham et al. 2014). It is said that the recurrence among the neuron’s connections makes the RNNs to present temporal relationship among the inputs, considered a type of memory.

In this work, we propose and evaluate ConvNets in the task of localization due to its inherent ability of feature extraction. Our proposed ConvNet topology is described in “Environment and methods for indoor localization” section.

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 the Received Signal Strength Indicator (RSSI) of wireless nodes. Figure 1 shows the environment: There are three routers (R1–R3) that broadcast RSSI signals and four rooms where collections were carried out (C1–C4). In each room (C1–C4), nine different collections were performed taking ≈ 3 to 4 minutes each (a total of 30 minutes in each room, and 2 hours taking into account the four rooms). Each scan of the networks takes approximately 0.4 seconds; hence, in 2 hours

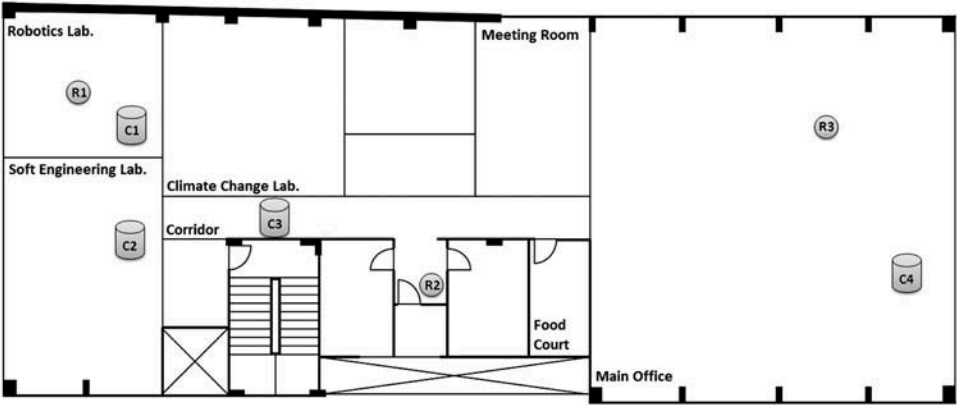


Figure 1. The testing environment: There are three routers (R1–R3) that broadcast RSSI signals and four rooms where collections were carried out (C1–C4). In each room (C1–C4), nine different collections were performed taking ≈ 3 to 4 minutes each (a total of 30 minutes in each room, and 2 hours taking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

of collection, the data set for the learning process of the ConvNet has a total of 17,617 records from signal strength. The data set 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 the 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.

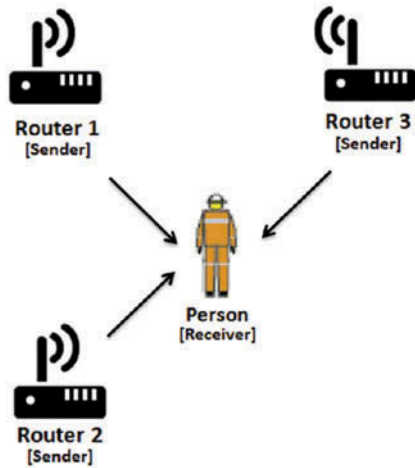


Figure 2. Architecture of the system: A person must have a device that receives RSSI from routers and employs machine learning to estimate its location.

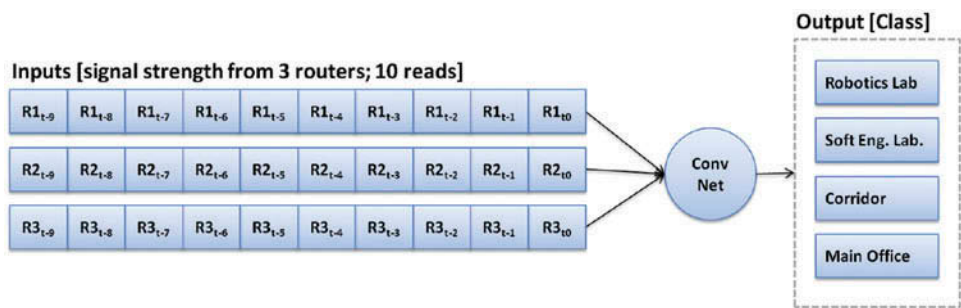


Figure 3. Simplified model of the inputs and outputs of the ConvNet. The ConvNet receives signal strengths as inputs, and it is trained to output the class related to each room of interest.

Related to the application of the ConvNet, 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 built to deal with time series, and is based on the work by Zheng and colleagues Zheng et al. (2014). The initial architecture is represented in Figure 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. We employed Rectified Linear Units (RELU) because it speeds up the training over classical activation functions Krizhevsky, Sutskever, and Hinton (2012) as sigmoid and hyperbolic tangent was used Dropout to increase the spread of the network and avoid overfitting Srivastava et al. (2014).

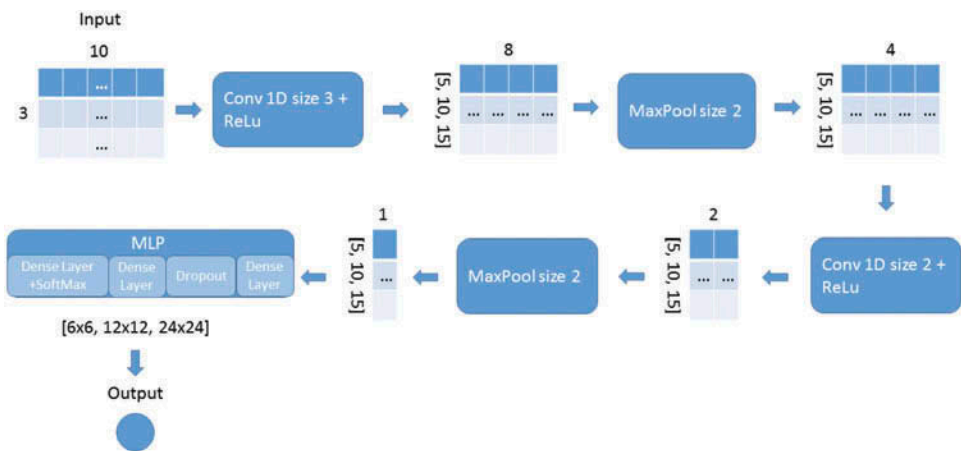


Figure 4. Convolutional Neural Network architecture, displaying the layers and feature maps generated. Between each feature map is shown a layer with the size of the filter. For each of the feature maps, brackets represent the number of filters in the convolution layer and the number of neurons in each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

The learning rate used was 0.001 and the momentum was 0.005 using Nesterov Momentum Sutskever et al. (2013). The number of trainable parameters ranged from approximately 1700 to 2000. All the network weights were 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 6×6 , 12×12 , 24×24 neurons in the MLP classification layer. Each network was trained with 20,000 epochs and the result aggregates 10 runs of each architecture.

Results

Figure 5 shows the results for the different ConvNet architectures. We can see that architectures with 5 filters in any convolution layer or 6×6 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 12×12 neurons each. Taking this into account, the architecture with 10 filters on both convolutional layers and 12×12 neurons in each hidden layer (c10c10d12) was considered the most appropriate for the task of classifying since it is the smallest architecture presenting best classification rates.

Figure 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 (Figure 1), which represent two neighbor rooms (Robotics Lab. and Soft Eng. Lab.).

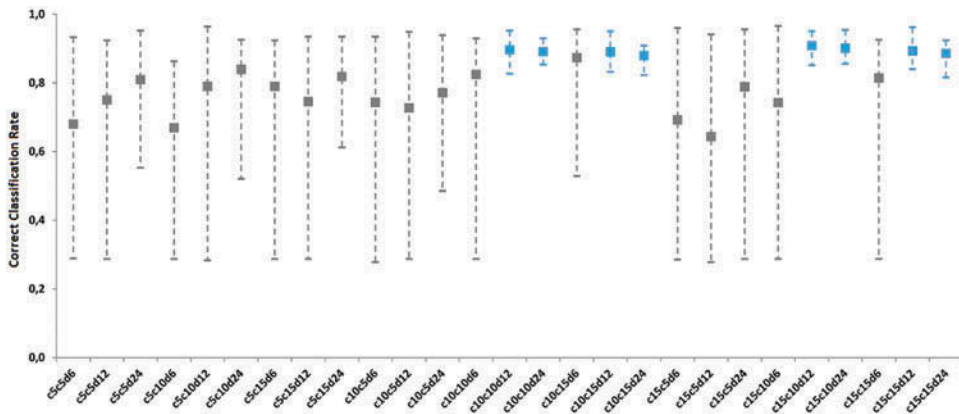


Figure 5. Results of correct classification rates for the evaluated ConvNet topologies. Each line presents the average of ten runs plus the max and min of the set. In blue, we show the best results, i.e. the ConvNets that presented correct classification rates with lower dispersion. All the best sets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

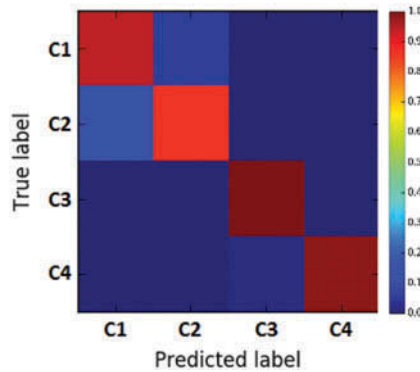


Figure 6. Confusion matrix of the best run of the architecture c10c10d12.

Conclusion and future work

In this paper, we exploited the use of ConvNets to perform localization of people in indoor environments. Several architectures were evaluated, seeking to understand which topology could obtain best results. As seen in this paper, ConvNets 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 to perform classification. The different architectures with several different layers showed that there are topologies more suitable to solve the problem.

Future studies should address other Deep Learning concepts by exploiting RNNs 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 which sensor (or sensors) could provide the best set to improve the accuracy of the system is a field that deserves attention.

Acknowledgments

The authors would like to acknowledge the following colleagues: Geraldo Pereira (USP), Bruno Faíçal (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) for their inspirational thoughts and several aids to the project. Jó Ueyama would like to thank FAPESP for funding the bulk of his research project.

Funding

The authors acknowledge the financial support from the “Edital SENAI SESI de Inovação (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., L. Deng, and D. Yu. 2013. Exploring convolutional neural network structures and optimization techniques for speech recognition. Proceedings of the 14th Annual Conference of the International Speech Communication Association (INTERSPEECH'13). Lyon, France, No. August, 3366–70.
- Bâce, M., and Y. A. Pignolet. 2015. Lightweight indoor localization system. Proceedings of the 8th IFIP Wireless and Mobile Networking Conference (WMNC), Munich, Germany, 160–167.
- Carvalho, E., B. S. Façal, G. P. R. Filho, P. A. Vargas, J. Ueyama, and G. Pessin. 2016. Exploiting the use of machine learning in two different sensor network architectures for indoor localization. Proceedings of the IEEE International Conference on Industrial Technology (ICIT), Taipei, Taiwan, 652–57.
- Dalto, M. 2015. Deep neural networks for time series prediction with applications in ultra-short-term wind forecasting. Proceedings of the IEEE International Conference on Industrial Technology (ICIT), Seville, Spain, 1657–63. <http://ieeexplore.ieee.org/xpl/articleDetails.jsp?arnumber=7125335>.
- Elnahrawy, E., X. Li, and R. P. Martin. 2004. The limits of localization using signal strength: A comparative study. Proceedings of the 1st Annual IEEE Communications Society Conference on Sensor and Ad Hoc Communications and Networks (IEEE SECON'04), Santa Clara, California, 406–14.
- Graves, A., and N. Jaitly. 2014. Towards end-to-end speech recognition with recurrent neural networks. Proceedings of the ICML, vol. 14, Beijing, China, 1764–72.
- Harris, J., P. Kirsch, M. Shi, J. Li, A. Gagrani, E. S. Krishna, A. Tabish, D. Arora, K. Kothandaraman, D. Cliff, Others. 2014. Comparative analysis of coal fatalities in Australia, South Africa, India, China and USA, 2006-2010. Proceedings of the 14th Coal Operator's Conference, Wollongong, Australia.
- Krizhevsky, A., I. Sutskever, and G. E. Hinton. 2012. ImageNet classification with deep Convolutional Neural Networks. Advances in Neural Information Processing Systems, Curran Associates, Inc., Lake Tahoe, Nevada, 1097–1105.
- Ladd, A. M., K. E. Bekris, A. P. Rudys, D. S. Wallach, and L. E. Kavraki. 2004. On the feasibility of using wireless ethernet for indoor localization. *IEEE Transactions on Robotics and Automation* 20 (3):555–59, June. doi:10.1109/TRA.2004.824948.
- LeCun, Y., Y. Bengio, and G. Hinton. 2015. Deep learning. *Nature* 521 (7553):436–44. doi:10.1038/nature14539.
- Lecun, Y., L. Bottou, Y. Bengio, and P. Haffner. 1998. Gradient-based learning applied to document recognition. *Proceedings of the IEEE* 86 (11):2278–324, November. doi:10.1109/5.726791.
- Pessin, G., F. S. Osório, J. Ueyama, D. F. Wolf, R. C. Moioli, and P. A. Vargas. 2014. Self-localisation in indoor environments combining learning and evolution with wireless networks. Proceedings of the 29th Annual ACM Symposium on Applied Computing, Gyeongju, Korea, 661–66. ACM.
- Pham, V., T. Bluche, C. Kermorvant, and J. Louradour. 2014. Dropout improves recurrent neural networks for handwriting recognition. Proceedings of the 14th International Conference on Frontiers in Handwriting Recognition (ICFHR), Crete, Greece, 285–90. IEEE.
- Sak, H., A. W. Senior, and F. Beaufays. 2014. Long short-term memory recurrent neural network architectures for large scale acoustic modeling. Proceedings of the INTERSPEECH, Singapore, 338–42.

- Sermanet, P., D. Eigen, X. Zhang, M. Mathieu, R. Fergus, and Y. LeCun. 2013. OverFeat: Integrated recognition, localization and detection using convolutional networks. arXiv preprint arXiv, 1312. 6229. <http://arxiv.org/abs/1312.6229>.
- Srivastava, N., G. E. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov. 2014. Dropout: A simple way to prevent neural networks from overfitting. *Journal of Machine Learning Research* 15:1929–58.
- Sutskever, I., J. Martens, G. E. Dahl, and G. E. Hinton. 2013. On the importance of initialization and momentum in deep learning. *JMLR W&Cp* 28 (2010):1139–47. <http://dblp.uni-trier.de/db/conf/icml/icml2013.html#SutskeverMDH13>.
- US Fire Administration. U.S. fire statistics: Trends in fires, deaths, injuries and dollar loss 2014. www.usfa.fema.gov/data/statistics/#tab-3.
- Yoo, J., T. Kim, C. Provencher, and T. Fong. 2014. WiFi localization on the international space station. Proceedings of the IEEE Symposium on Intelligent Embedded Systems (IES), Orlando, FL, USA, 21–26. IEEE.
- Zheng, Y., Q. Liu, E. Chen, Y. Ge, and J. L. Zhao. 2014. Time series classification using multi-channels deep convolutional neural networks. Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), Macau, China, vol. 8485, 298–310. LNCS. doi:[10.1007/978-3-319-08010-9_33](https://doi.org/10.1007/978-3-319-08010-9_33).