Development of a highly precise place recognition module for effective human-robot interactions in changing lighting and viewpoint conditions

Hermann Baumgartl
Aalen University, Germany
hermann.baumgartl@hs-aalen.de

Ricardo Buettner
Aalen University, Germany
ricardo.buettner@hs-aalen.de

Abstract

We present a highly precise and robust module for indoor place recognition, extending the work by Lemaignan et al. and Robert Jr. by giving the robot the ability to recognize its environment context. We developed a full end-to-end convolutional neural network architecture, using a pre-trained deep convolutional neural network and the explicit inductive bias transfer learning strategy. Experimental results based on the York University and Rzeszów University dataset show excellent performance values (over 94.75 and 97.95 percent accuracy) and a high level of robustness over changes in camera viewpoint and lighting conditions, outperforming current benchmarks. Furthermore, our architecture is 82.46 percent smaller than the current benchmark, making our module suitable for embedding into mobile robots and easily adoptable to other datasets without the need for heavy adjustments.

1. Introduction

While mobile robots have been used in industrial environments for decades, they became affordable for the consumer market in recent years as well. In many homes, domestic robots, self-driving cars and voice-activated assistants and interactions with them are common nowadays. One of the most important aspects of Human-Robot-Interaction (HRI) is the robots ability to understand abstract spatial concepts and act accordingly [1, 2]. For example, a robot vacuum cleaner may be asked to clean the bedroom, therefore the robot's recognition of a bedroom should match the human's understanding of such a place [1].

Interaction with robots is becoming an important part in modern work environments [2, 3]. Human-robot teamwork is one of the most important topics of HRI research [2, 4]. Studies on how to promote performance in human-robot teamwork have shown the positive effects of trust [4, 5, 6], team diversity [3], emotional attachment [7] and robot personality [8]. However the impact of context on robot personality [8] and contextual decision making [2] is still open. To overcome the lack of contextual reasoning, Lemaignan et al. [2] proposed an architecture combining the Belief-Desires-Intention (BDI) architecture and cognitive skills into the robot.

Research has shown that not only designing the robot to increase a user's situational awareness, but also to implementation it into the robot itself (see Fig. 1), improves the robot's decision making [9]. The theory of situational awareness [10] by Endsley consists of three consecutive levels: perception of the elements in the situation (level 1), comprehension of the situation (level 2) and projection of future actions (level 3). A weakness of most mobile robots is their lack of level 1 capabilities. Robots only have limited awareness of the environment they are in and therefore they can only be partially capable of contextual reasoning [2].

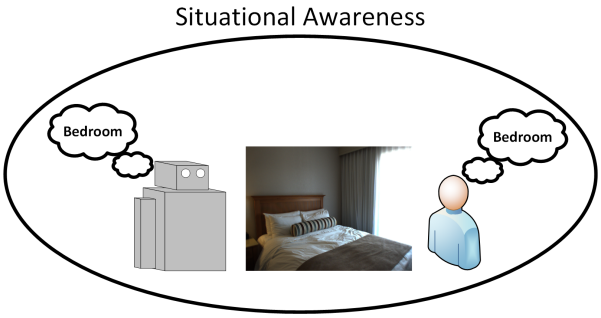

Figure 1. Situational awareness for context aware mobile robots.

That is why we developed a robust place recognition module for robot localization, to substantially enhance a robot's level 1 capabilities. Robot localization and navigation has been extensively studied, using Lidar [11], Sonar [12], GPS [13] and sensor information [14]. Place recognition on the other hand is mostly based on categorizing monocular images into predefined classes [15]. Robust place recognition is one of the most difficult challenges, due to the high complexity of the places themselves [16], changing lighting and viewpoint

URI: https://hdl.handle.net/10125/63808
978-0-9981331-3-3

condition [15], occlusions of objects and landmarks [16] and the presence of other dynamic objects [17].

We use the MobileNetV2 [18] architecture – a highly optimized convolutional neural networks developed for mobile and embedded devices – to build our place recognition module. We train and evaluate our module using the York University 11 and 17 places [15] and the Rzeszów University 16 places dataset [19], outperforming current benchmarks. Since both datasets are relatively small we use a transfer learning approach based on the explicit inductive bias L^2-*SP* [20]. Our results demonstrate the robustness of our module and the advantage of a full end-to-end convolutional neural network over its sole use as feature extractor.

In order to improve our previous approaches [9, 21, 22], we propose a module which can be implemented into embedded devices such as mobile robots, shows robust classification accuracy in various environments and can be easily adopted to other environments and datasets.
Our most important contributions are:

1. We developed a highly precise indoor place recognition module which outperforms the current York University and Rzeszów University dataset benchmarks with an accuracy of over 94.75 and 97.95 percent [19].
2. Our module is 82.46 percent smaller than the benchmark models and therefore needs significantly less computational power [19].
3. We extend the work of Lemaignan et al. [2] and Robert Jr. [8] by giving the robot the ability to precisely recognize its environment context.
4. Our approach is robust against severe changes in lighting and viewpoint conditions [1] and shows good performance in different environments [15].
5. Our module closes the gap between visual level 1 situational awareness capabilities for mobile robots and HRI [8].
6. Our module can be easily adopted for other mobile robot scenarios, without the need to massively change the networks' configuration.

The paper is organized as follows: First we give an overview of the related work, including a description of the current approaches for place recognition. Next we provide the research methodology, with a description of the deep learning methods and data used. After that we show results of our recognition modules performance and compare it with current benchmarks. We then discuss the results and its implications, before concluding with limitations and suggestions for future research.

2. Related Work

Robust place recognition for mobile robots is one of the major challenges of HRI research. The high complexity of robust place recognition comes from changes in the scene [16], varying lighting and viewpoint conditions [23], the limited computational resources [18] and the high complexity of the places [1].

Earlier work on place recognition based on camera images mostly adopted two separate stages: 1) extract hand crafted feature from the images, 2) use a (shallow) machine learning model like Support Vector Machine to classify the image into a predefined category. Since most shallow learning algorithms perform poorly on raw image data, both global and local image descriptors have been used to extract image features [23, 24, 25, 26, 27]. These feature descriptors extract textural features such as edges or bright and dark spots, which can be used to categorize the image [15]. Due to the high variability of environments conditions, handcrafted descriptors struggle with robustness [1, 17].

More recently, feature extractors based on convolutional neural network have gained more attention, substantially improving the quality of robot place recognition. The results from [1, 16, 17, 19] showed that convolutional neural networks are more robust against changing image conditions than handcrafted descriptors. However these approaches are mostly based on very large ImageNet-winning network architectures, which are very powerful but also very computationally expensive at the same time [18, 28]. Most algorithms perform well under constant conditions, but experience a significant drop in classification accuracy under changing conditions. A highly precise, robust and computationally efficient solution for place recognition is still missing [1].

3. Methodology

In this section we describe the use of deep learning methods and transfer learning strategies. To rigorously evaluate our module we follow the specific machine learning guidelines [18, 29, 30, 31] and conduct a comprehensive literature research [32, 33, 34].

3.1. Convolutional neural networks

In recent years convolutional neural networks (re-)emerged as the top method for solving complex computer vision tasks. One of the major advantages of convolutional neural network is the combination of feature extraction and final classification into one step. Using the backpropagation algorithm, the convolutional layers, i.e. the feature extraction, is optimized with

respect to the final classification task. Therefore convolutional neural networks are able to automatically learn a set of deep feature representation for high quantities of different objects [31].

3.2. Transfer learning strategies

Compared to traditional machine learning methods, deep convolutional neural networks are highly dependent upon massive amounts of data for training. Usually 100,000 images or more are needed to train a deep convolutional Network from scratch. In most application scenarios the acquisition of such amounts of data is unfeasible for various reasons [35]. However, the necessary amount of training data can be greatly reduced using transfer learning. Transfer learning tries to transfer learned representation from another domain to the target domain [20, 36].

The explicit inductive bias is a new strategy for transfer learning with convolutional neural networks, proposed by Li et al. [20]. In order to preserve the initial knowledge of the network, an adapted L^2 weight regularization is used. The so-called *L^2-SP* regularization prohibits the network from changing the networks weights too far from initial pre-trained weights. The degree of regularization can be adjusted using the α and β parameters, where higher values carry greater penalties [20].

3.3. Embedded convolutional neural network architectures

In the search for more and more powerful convolutional neural networks, researchers started to massively increase network sizes to generate more powerful networks. Most of the current high-performance convolutional neural networks have very high amounts of network parameters and therefore carry high computational cost [18]. Networks such as VGG16 (138.3M parameters) [37] or ResNet-50 (25.5M parameters) [38] exceed the hardware capabilities of most mobile or embedded devices and are therefore not feasible for implementation.

Since computational and energy resources are limited on embedded devices, convolutional neural networks can only be implemented using efficient and lightweight architectures [39]. To build lightweight networks different approaches, such as training small networks from scratch [9, 40] or shrinking pre-trained networks [41, 42] were developed. Training small networks from scratch has been successfully used to build very small and efficient networks [9, 40], but architecture engineering and optimization steps require significant amounts of time [43].

In addition to the existing pre-trained networks, architectures specifically built for implementation in mobile and embedded devices have been proposed in the literature. Networks like MobileNetV2 [18], NASNet [43] and ShuffleNet [28] provide efficient network architectures for applications with low computational resources.

3.4. MobileNetV2 architecture

The MobileNetV2 architecture proposed by Sandler et al. [18] is a convolutional neural network architecture optimized to provide low memory consumption and low computational cost during inference. Inference refers to running the model on new data to generate predictions. The main improvement of MobileNetV2 is the introduction of the inverted-residuals-with-linear-bottleneck-blocks, which greatly reduced the memory consumption. The size of MobileNetV2 can be adjusted using the width multiplier and the input size. The width multiplier thins or widens the network at each layer, smaller input sizes such as 96x96 pixels reduce the networks' size and computational costs in exchange for lower classification accuracy. The standard MobileNetV2 network uses a width multiplier of 1.0 and 224x224 input images [18]. Besides the computational efficiency of the MobileNetV2 architecture, one of its main benefits is the broad availability of pre-trained MobileNetV2 implementation in many common deep learning frameworks.

3.5. Evaluation data and data pre-processing

We trained and evaluated our recognition module on the York University 11 and 17 places and the Rzeszów University 16 places datasets [15, 19]. The first dataset consists of 11 indoor places captured by two robots (Pioneer and Virtual Me) under different lighting conditions (daytime and nighttime) at York University. The images were acquired using a color camera mounted on top of the robots. For the *Pioneer* robot the camera is 88 centimeters above the ground, whereas the camera on the *Virtual Me* robot is 117 centimeters above the floor. For the 17 places dataset, images were only acquired using the *Virtual Me* robot, representing six additional places at the Coast Capri Hotel, Kelowna in British Columbia [15]. As in previous work by [15, 19] the place recognition accuracies are tested in four scenarios:

1. Same robot for training and testing with same lighting conditions
2. Same robot for training and testing with different lighting conditions

3. Different robots for training and testing with same lighting conditions

4. Different robots for training and testing with different lighting conditions

The images were resized to 224x224 pixels and scaled to [-1,1]. The total dataset size was 13,751 RGB images for the 11 places dataset and 16,110 images for the 17 places dataset. To improve the classification accuracy we used data augmentation to enhance the training data, by rotating the images 90 and 270 degrees, horizontal and vertical flipping and applying random image noise and blur. Data augmentation is a common technique used to increase classification performance and decrease model overfitting [44]. Examples of each of the original images are given in Fig. 2.

Figure 2. Sample images of the 17 places from [15].

In order to further evaluate our modules' generalization capability on a greater variety of possible environments, we used the Rzeszów University dataset [19]. This dataset contains 8,000 RGB images of 16 different indoor places capture by the Nao humanoid robot at the Rzeszów University of Technology. The images were also resized to 224x224 pixels, scaled to [-1,1] and enhanced using the data augmentation used on the York University datasets. Unlike the York University dataset, the Rzeszów University dataset only contains images from a single lighting condition.

For performance evaluation of the predictor we use a *hold-out cross validation*. The hold-out cross validation splits the dataset into training and testing subsets. The evaluation is only done using the testing data. This data must not be shown to the model before evaluation. Potential overfitting of the model could be identified during the evaluation [30].

4. Results

To train the convolutional neural network we used the Keras 2.1.5 package [45] with TensorFlow 1.8 backend [46]. The training for all datasets ran on a Nvidia GeForce GTX 1080 Ti for 20 epochs, using transfer learning with explicit inductive bias ($\alpha = 0.1$, $\beta = 0.01$) and a RMSprop optimizer with an initial learning rate of 1e-5. The evaluation data has not been shown to the model during training. Execution speed for inference was tested on an Intel Core i7-8750H notebook CPU and a Nvidia GeForce GTX 1050 Ti notebook GPU.

4.1. Our recognition module architecture

Our recognition module utilizes a transfer learning approach based on MobileNetV2 with 1.0 width-multiplier, 224x224 pixels input size and explicit inductive bias as the transfer learning strategy. The final classification layer at the end of the MobileNetV2 model has been replaced with a 11-node, 16-node or 17-node fully-connected layer, depending on the dataset used. Besides changing the output layer to match the number of classes in each dataset, no further adjustments to the network have been made. The model receives RGB images shaped 224x224x3. The final model contained 2.58M weights. The distribution of weights across the blocks is shown in table 1.

4.2. Performance evaluation

For the York University 11-places dataset, we compared our model with two state-of-the-art approaches by Sahdev et al. [15] and Wozniak et al. [19]. As shown in table 2 our model delivers a state-of-the-art

Input	Layer	Rep.	Weights
224x224x3	conv2d	1	992
112x112x32	inv_bottleneck	1	705
112x112x16	inv_bottleneck	1	5,568
56x56x24	inv_bottleneck	1	9,456
56x56x144	inv_bottleneck	1	10,512
28x28x32	inv_bottleneck	3	53,312
14x14x64	inv_bottleneck	4	236,169
14x14x96	inv_bottleneck	3	399,040
7x7x160	inv_bottleneck	3	1,126,720
7x7x1280	conv2d 1x1	1	414,720
7x7x1280	avgpooling2d	1	0
1x1x1280	dense-256	1	327,936
1x1x256	output	1	2,827

Table 1. Architecture of our module for 11-places. Kernel size for all spatial convolutions is 3x3. Weights are the total amount of weights per block. Rep. denotes the times each block is repeated.

performance. In particular, even for the most challenging experiment IV, our model achieves classification accuracies above 90 percent, outperforming all the baselines.

In the first and second experiments our module shows comparable results with the benchmarks. The high classification accuracy shows its robustness against changing lighting conditions, while experiment III demonstrates robustness against changes in camera position on the robot platform. In particular in the experiment IV, where both heavy changes in viewpoint and lighting conditions occur, our module shows good improvements in classification accuracy above the current benchmarks. This shows our module's ability to generalize, irrespective of changing input conditions.

We further evaluated the performance of our recognition module in the most challenging experiment IV, with Virtual Me as training set and Pioneer as testing set under different lighting conditions. Evaluation metrics are class-averaged sensitivity (true positive rate), precision (positive predictive value), Cohen's Kappa score and accuracy. As shown in table 3, the classifier achieved excellent performance values.

The high classification accuracy under different lighting conditions and with different robots show the clear advantage of convolutional neural network based place recognitions modules over traditional and mixed methods. When comparing our recognition module on the 17 place dataset in table 4, our module again shows better overall classification accuracy especially under changing lighting conditions. While the original 11 place dataset all represent location at York University, the 17 places dataset also contains images taken at the Coast Capri Hotel, Kelowna in British Columbia [15]. The 17 places dataset represents a larger variety of places and increases the complexity of the recognition task. The results of the second experiment show a good improvement in classification accuracy over the current benchmark. This suggests that our module is capable of adapting to different locations and place appearances, while still maintaining robustness across changing lighting conditions.

To further evaluate our modules generalization capability on different environments, we compared our modules performance on the Rzeszów University dataset for 16 places. As shown in table 5 our recognition module achieves a very good classification result of 97.95 percent, again outperforming the previous results [19]. The results of the execution speed evaluation is shown in table 6. The evaluation was done using different batch sizes. Since graphic chips in particular are highly optimized for parallel processing, larger batch sizes lead to higher performance on runtime. In line with the original result by Sandler et al. [18] the MobileNetV2 base of our recognition module provides good computational speed even on the CPU. Increasing the batch size on the CPU nearly linear increases the computation time needed.

5. Discussion

As shown in table 2 our model is capable of robustly identifying the robot's location under changing lighting and viewpoint conditions. With a mean accuracy of 94.75 percent over the York University experiments, our recognition module outperforms all current benchmarks. Additionally in the toughest experiment IV, our model shows its robustness against changing lightning conditions and heavy viewpoint changes. Results on the 17 places dataset further underpins this, by showing the module's ability to adopt to other locations as well.

The results of the performance comparison in table 3 and table 4 show the benefit of using an end-to-end convolutional neural network architecture. The combination of a convolutional neural network for feature extraction feeding into a Support Vector Machine for classification in [19], shows high accuracy for static lighting and viewpoint conditions. However, performance drops under a combination of changing lighting and viewpoint conditions. Our proposed module achieves very high recognition accuracies and maintains robustness against changing conditions, while still being low on computational costs. The further evaluation on the Rzeszów University dataset in table 5 shows the ability of our module to adapt towards different

Experiment	Training Set	Testing Set	Lighting	Accuracy (%) Sahdev et al. [15]	Wozniak et al. [19]	**Ours**
I	Pioneer	Pioneer	same	98	99	98.91
	Virtual Me	Virtual Me	same	98	98	98.59
II	Pioneer	Pioneer	different	93	94	92.57
	Virtual Me	Virtual Me	different	93	92	93.97
III	Pioneer	Virtual Me	same	92	92	94.02
	Virtual Me	Pioneer	same	92	95	96.38
IV	Pioneer	Virtual Me	different	82	86	92.13
	Virtual Me	Pioneer	different	85	89	91.48
			Mean	91.63	93.13	**94.75**

Table 2. Comparison of recognition accuracy on the York University dataset [15] for eleven places.

Performance indicator	Value (%)
Accuracy	91.090
True positive rate	91.487
True negative rate	99.092
Positive predictive value	91.033
Negative predictive value	99.085
Prevalence	9.091
Balanced accuracy	91.487
Kappa	90.019

Table 3. Evaluation indicators of our object recognition module.

#	Training	Testing	Lighting	Accuracy (%) [15]	**Ours**
I	VME	VME	same	98.34	98.60
II	VME	VME	different	90.22	93.97
			Mean	94.28	**96.29**

Table 4. Recognition accuracy on the York University dataset [15] for seventeen places. VME denotes the Virtual Me robot platform.

environments. While the Rzeszów University dataset also mostly contains typical office environments, the results show the very high adaption capability of our module with limited data and in environments with different shapes or configurations of the location itself.

The evaluation of the inference performance demonstrates the high execution speed of our module. The computations took 72ms per image on a CPU, with a batch-size of 1. Using larger batch sizes of 16 images per batch, almost linearly increases the computational time needed. Since convolutional neural networks mostly benefit from parallel processing, increases in clock speed of the CPU only have a marginal impact on the performance of the network [47]. However, as the comparison of computational speed shows, GPU powered model inference greatly increases the efficiency of the network. The usage of a GPU powered embedded platform such as the Nvidia Jetson Nano, could enable model inference in real-time. The total of 2.58M weights is considerably less than for the VGG-F model (14.71M for the convolutional base only) from [19], resulting in 82.46 percent size reduction. This large reduction in size enables our module to be integrated into the mobile robot itself, whereas the much larger VGG-F network consumes high amounts of computational resources and is therefore not feasible for devices with limited computational power.

6. Conclusion

Using L^2-*SP* implementation of the explicit inductive bias as a transfer learning strategy with the MobileNetV2 architecture as base network, we developed a highly effective indoor robot localization module, showing state-of-the-art results while being robust against changes in image lighting and camera viewpoint. The performance evaluation showed 91.487 percent balance accuracy for the most complex evaluation scenario with changing lighting and viewpoint conditions on the York University dataset. The additional evaluation of the Rzeszów University dataset shows our modules' ability to adapt towards different environments. In contrast to the approach by [19], our module did not require any parameter optimization to achieve good classification results when applied to different datasets. Since our module is based on commonly available open source software and MobileNetV2, other scholars can easily adapt our architecture. Our module achieved very good results with limited data after a short training cycle of 20 epochs, while the combination of heavy data augmentation and L^2-*SP* effectively reduced overfitting.

	Accuracy (%)	Precision	Recall	F1-score
BoW, SURF, SVM [19]	79.07	79.07	73.12	73.10
VGG-F, SVM [19]	95.13	95.10	94.83	94.88
VGG-F, SVM, no-blur [19]	95.44	95.44	95.49	95.36
VGG-F fine-tuned [19]	97.19	97.12	97.20	97.16
VGG-F fine-tuned, aug. [19]	96.69	96.57	96.76	96.66
OURS	**97.95**	**97.97**	**97.95**	**97.95**

Table 5. Recognition accuracy on the Rzeszów University dataset [19] for sixteen places.

Batch Size	CPU	GPU
1	72ms	11ms
2	128ms	15ms
8	523ms	37ms
16	1117ms	67ms
25	1848ms	99ms
32	2447ms	126ms

Table 6. Execution speed for inference on CPU and GPU using different batch sizes.

6.1. Limitations

One limitation is related to the datasets. While they provide data for different places, these places are mostly located at York or Rzeszów universities. While our module shows good performance on both the 11 and 17 place York University dataset as well as the Rzeszów University dataset, a majority of the images represent typical office and laboratory environments. The 17 places dataset adds six more locations from a hotel environment, which partially displays the model's the generalization ability to adapt to a completely different environment. However, common domestic places like living rooms are absent in the datasets. Furthermore these images are available for the Virtual Me robot platform only. Tests with more datasets representing a greater variety of places are necessary to fully evaluate the robustness of our module and its generalization capabilities. Furthermore, object occlusion by other moving objects, such as humans moving around, are only present in a minor portion of the York University 17 places dataset. Since in real world scenarios moving objects are common, these factors also have to be tested in the future.

Another limitation is related to the hardware used for performance evaluation. Calculations are currently done on a standard laptop, however this laptop exceeds the hardware capabilities of most embedded platforms. While training is typically done using a powerful workstation or laptop, inference of the trained module takes place on the embedded hardware. More tests, using different embedded platforms such as Raspberry Pi or Nvidia Jetson Nano are necessary to evaluate the performance of our module on embedded devices.

6.2. Future work

In order to fully asses our module's generalization capability, we will re-evaluate its performance using more datasets with a much greater variety of places. We will re-evaluate our module using datasets like SUN397 [48], MIT Indoor67 [49], KTH-IDOL2 [50] and Scene15 [51]. These datasets represent a broad variety of different indoor and outdoor environments and also include occlusion by moving objects. Furthermore we will provide inference runtime comparisons for different embedded platforms.

Additional challenges for mobile robots not only arise from place recognition, but also from object detection. Results by Sandler et al. [18] show that the MobileNetV2 architecture is a viable base for object detection tasks as well. In future research we will use our existing MobileNetV2 base for place recognition and object detection tasks in a combined module.

Another future research line which we will follow is to investigate how a user's cognitive workload and related user-oriented concepts [52, 53, 54] change in real-world Human-Robot-Interactions due to the place recognition module. Therefore we plan experiments

- to assess mental concepts such as cognitive workload [55, 56, 57], concentration [58], and mindfulness [59, 60] when using our place recognition module in real-word Human-Robot-Interactions in multi-agent-settings [61, 62, 63, 64],
- to triangulate objective and perceived user-oriented concepts [65, 66, 67] using physiological sensor data (i.e., electroencephalographic data [68, 69, 70, 71] and spectra [72, 73, 74], electrocardiographic data [75, 76], electrodermal activity [77], eye fixation [78, 79, 56], eye pupil diameter [80, 81, 53, 82], facial expressions [83]), and,

- to evaluate technology acceptance [84, 85, 86, 87] and trust [88] of our embodied module and confirm if the automated approach improves the coordination [89, 90, 91, 92, 93, 94, 95] more efficiently.

To further improve our models performance, we will pre-train the MobileNetV2 architecture on a scene recognition specific dataset like Places365 [96], therefore providing a domain specific convolutional base for place recognition tasks.

Acknowledgments

We would like to thank the reviewers, who provided very helpful comments on the refinement of the HICSS paper. This research is partly funded by the German Federal Ministry of Education and Research (no. 13FH4E03IA, no. 13FH4E07IA, no. 13FH176PX8).

References

[1] M. Mancini, S. R. Bulo, E. Ricci, and B. Caputo, "Learning Deep NBNN Representations for Robust Place Categorization," *IEEE Robot. Autom. Lett.*, vol. 2, no. 3, pp. 1794–1801, 2017.

[2] S. Lemaignan, M. Warnier, E. A. Sisbot, A. Clodic, and R. Alami, "Artificial cognition for social human–robot interaction: An implementation," *Artif. Intell.*, vol. 247, pp. 45–69, 2017.

[3] S. You and L. P. Robert Jr., "Team Potency and Ethnic Diversity in Embodied Physical Action (EPA) Robot-Supported Dyadic Teams," in *ICIS 2017 Proc.*, AIS, 2017.

[4] S. You and L. P. Robert Jr., "Trusting Robots in Teams: Examining the Impacts of Trusting Robots on Team Performance and Satisfaction," in *HICSS-52 Proc.*, pp. 244–253, IEEE, 2019.

[5] S. You, J.-H. Kim, S. Lee, V. Kamat, and L. P. Robert, "Enhancing perceived safety in human-robot collaborative construction using immersive virtual environments," *Automat. Constr.*, vol. 96, pp. 161–170, 2018.

[6] S. You and L. P. Robert Jr., "Human-Robot Similarity and Willingness to Work with a Robotic Co-worker," in *HRI '18 Proc.*, ACM, 2018.

[7] S. You and L. P. Robert Jr., "Emotional Attachment, Performance, and Viability in Teams Collaborating with Embodied Physical Action (EPA) Robots," *J. Assoc. Inf. Syst.*, vol. 19, no. 5, pp. 377–407, 2018.

[8] L. P. Robert Jr., "Personality in the Human Robot Interaction Literature: A Review and Brief Critique," in *AMCIS 2018 Proc.*, AIS, 2018.

[9] R. Buettner and H. Baumgartl, "A Highly Effective Deep Learning Based Escape Route Recognition Module for Autonomous Robots in Crisis and Emergency Situations," in *HICSS-52 Proc.*, pp. 659–666, IEEE, 2019.

[10] M. R. Endsley, "Toward a Theory of Situation Awareness in Dynamic Systems," *Human Factors*, vol. 37, no. 1, pp. 32–64, 1995.

[11] J. Zhang and S. Singh, "Visual-lidar Odometry and Mapping: Low-drift, Robust, and Fast," in *ICRA 2015 Proc.*, IEEE, 2015.

[12] A. Sanchez, A. d. Castro, S. Elvira, G. Glez-de Rivera, and J. Garrido, "Autonomous indoor ultrasonic positioning system based on a low-cost conditioning circuit," *Measurement*, vol. 45, no. 3, pp. 276–283, 2012.

[13] M. Agrawal and K. Konolige, "Real-time localization in outdoor environments using stereo vision and inexpensive GPS," in *ICPR '06 Proc.*, IEEE, 2006.

[14] O. Woodman and R. Harle, "Pedestrian Localisation for Indoor Environments," in *UbiComp '08 Proc.*, 2008.

[15] R. Sahdev and J. K. Tsotsos, "Indoor Place Recognition System for Localization of Mobile Robots," in *CRV 2016 Proc.*, IEEE, 2016.

[16] R. Arroyo, P. F. Alcantarilla, L. M. Bergasa, and E. Romera, "Fusion and Binarization of CNN Features for Robust Topological Localization across Seasons," in *IROS 2016 Proc.*, IEEE, 2016.

[17] N. Suenderhauf, S. Shirazi, F. Dayoub, B. Upcroft, and M. Milford, "On the performance of ConvNet features for place recognition," in *IROS 2015 Proc.*, IEEE, 2015.

[18] M. Sandler, A. Howard, M. Zhu, A. Zhmoginov, and L.-C. Chen, "MobileNetV2: Inverted residuals and linear bottlenecks," in *CVPR 2018 Proc.*, IEEE, 2018.

[19] P. Wozniak, H. Afrisal, R. G. Esparza, and B. Kwolek, "Scene recognition for indoor localization of mobile robots using deep CNN," in *ICCVG 2018 Proc.*, pp. 137–147, Springer, 2018.

[20] X. Li, Y. Grandvalet, and F. Davoine, "Explicit Inductive Bias for Transfer Learning with Convolutional Networks," in *PMLR 2018 Proc.*, pp. 2825–2834, 2018.

[21] H. Baumgartl, R. Buettner, T. Bernthaler, I. J. Timm, A. Jansche, and G. Schneider, "Colored micrographs significantly outperform grayscale ones in modern machine learning: Insights from a systematical analysis of lithium-ion battery micrographs using convolutional neural networks," in *MCM 2017 Proc.*, pp. 98–101, 2017.

[22] H. Baumgartl, J. Tomas, R. Buettner, and M. Merkel, "A novel deep-learning approach for automated non-destructive testing in quality assurance based on convolutional neural networks.," in *ACEX 2019 Proc.*, 2019.

[23] A. Pronobis, B. Caputo, P. Jensfelt, and H. Christensen, "A Discriminative Approach to Robust Visual Place Recognition," in *IROS '06 Proc.*, IEEE, 2006.

[24] E. Fazl-Ersi and J. K. Tsotsos, "Histogram of Oriented Uniform Patterns for robust place recognition and categorization," *Int. J. Robot. Res.*, vol. 31, no. 4, pp. 468–483, 2012.

[25] A. Oliva and A. Torralba, "Modeling the Shape of the Scene: A Holistic Representation of the Spatial Envelope," *Int. J. Comput. Vis.*, vol. 42, no. 3, pp. 145–175, 2001.

[26] D. G. Lowe, "Distinctive Image Features from Scale-Invariant Keypoints," *Int. J. Comput. Vis.*, vol. 60, no. 2, pp. 91–110, 2004.

[27] H. Bay, T. Tuytelaars, and L. Van Gool, "SURF: Speeded up robust features," in *ECCV '06 Proc.*, pp. 404–417, Springer, 2006.

[28] X. Zhang, X. Zhou, M. Lin, and J. Sun, "ShuffleNet: An Extremely Efficient Convolutional Neural Network for Mobile Devices," in *CVPR 2018 Proc.*, IEEE, 2018.

[29] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, no. 7553, pp. 436–444, 2015.

[30] S. Arlot and A. Celisse, "A survey of cross-validation procedures for model selection," *Stat. Surv.*, vol. 4, pp. 40–79, 2010.

[31] J. Yosinski, J. Clune, Y. Bengio, and H. Lipson, "How Transferable Are Features in Deep Neural Networks?," in *NIPS '14 Proc.*, pp. 3320–3328, MIT, 2014.

[32] J. Webster and R. T. Watson, "Analyzing the past to prepare for the future: Writing a literature review," *MIS Quarterly*, vol. 26, no. 2, pp. xiii–xxiii, 2002.

[33] J. vom Brocke, A. Simons, B. Niehaves, K. Riemer, R. Plattfaut, and A. Cleven, "Reconstructing the Giant: On the Importance of Rigour in Documenting the Literature Search Process," in *ECIS '09 Proc.*, pp. 2206–2217, AIS, 2009.

[34] J. vom Brocke, A. Simons, K. Riemer, B. Niehaves, R. Plattfaut, and A. Cleven, "Standing on the Shoulders of Giants: Challenges and Recommendations of Literature Search in Information Systems Research," *Communications of the AIS*, vol. 37, 2015. Article 9.

[35] K. Weiss, T. M. Khoshgoftaar, and D. Wang, "A survey of transfer learning," *J. Big Data*, vol. 3, no. 1, 2016.

[36] S. J. Pan and Q. Yang, "A Survey on Transfer Learning," *IEEE T. Knowl. Data. En.*, vol. 22, no. 10, pp. 1345–1359, 2010.

[37] K. Simonyan, A. Vedaldi, and A. Zisserman, "Deep Fisher Networks for Large-Scale Image Classification," in *NIPS '13 Proc.*, pp. 163–171, 2013.

[38] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *CVPR 2016 Proc.*, IEEE, 2016.

[39] A. Canziani, E. Culurciello, and A. Paszke, "Evaluation of neural network architectures for embedded systems," in *ISCAS 2017 Proc.*, IEEE, 2017.

[40] S. Arnold and K. Yamazaki, "Real-time scene parsing by means of a convolutional neural network for mobile robots in disaster scenarios," in *ICIA 2017 Proc.*, 2017.

[41] J. Cheng, J. Wu, C. Leng, Y. Wang, and Q. Hu, "Quantized CNN: A Unified Approach to Accelerate and Compress Convolutional Networks," *IEEE Trans. Neural Netw. Learn. System.*, vol. 29, no. 10, pp. 4730–4743, 2018.

[42] J. Wu, C. Leng, Y. Wang, Q. Hu, and J. Cheng, "Quantized Convolutional Neural Networks for Mobile Devices," in *CVPR 2016 Proc.*, IEEE, 2016.

[43] B. Zoph, V. Vasudevan, J. Shlens, and Q. V. Le, "Learning Transferable Architectures for Scalable Image Recognition," in *CVPR 2018 Proc.*, IEEE, 2018.

[44] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet Classification with Deep Convolutional Neural Networks," in *NIPS '12 Proc.*, pp. 1097–1105, 2012.

[45] F. Chollet *et al.*, "Keras." `https://keras.io`, 2015.

[46] M. Abadi, P. Barham, J. Chen, Z. Chen, A. Davis, J. Dean, M. Devin, S. Ghemawat, G. Irving, M. Isard, M. Kudlur, J. Levenberg, R. Monga, S. Moore, D. G. Murray, B. Steiner, P. Tucker, V. Vasudevan, P. Warden, M. Wicke, Y. Yu, and X. Zheng, "Tensorflow: A system for large-scale machine learning," in *USENIX-OSDI 2016 Proc.*, pp. 265–283, 2016.

[47] J. Schmidhuber, "Deep learning in neural networks: An overview," *Neural Networks*, vol. 61, pp. 85–117, 2015.

[48] J. Xiao, K. A. Ehinger, J. Hays, A. Torralba, and A. Oliva, "SUN database: Exploring a large collection of scene categories," *Int. J. Comput. Vis.*, vol. 119, no. 1, pp. 3–22, 2014.

[49] A. Quattoni and A. Torralba, "Recognizing Indoor Scenes," in *CVPR 2009 Proc.*, pp. 413–421, IEEE, 2009.

[50] J. Luo, A. Pronobis, B. Caputo, and P. Jensfelt, "Incremental learning for place recognition in dynamic environments," in *IEEE/RSJ IROS 2007 Proc.*, 2007.

[51] F.-F. Li and P. Perona, "A Bayesian Hierarchical Model for Learning Natural Scene Categories," in *CVPR 2005 Proc.*, IEEE, 2005.

[52] R. Buettner, "Analyzing Mental Workload States on the Basis of the Pupillary Hippus," in *NeuroIS '14 Proc.*, p. 52, 2014.

[53] R. Buettner, "Investigation of the Relationship Between Visual Website Complexity and Users' Mental Workload: A NeuroIS Perspective," in *Information Systems and Neuro Science*, vol. 10 of *LNISO*, pp. 123–128, 2015.

[54] R. Buettner, S. Sauer, C. Maier, and A. Eckhardt, "Real-time Prediction of User Performance based on Pupillary Assessment via Eye Tracking," *AIS Trans. Hum.-Comput. Interact.*, vol. 10, no. 1, pp. 26–56, 2018.

[55] R. Buettner, "The relationship between visual website complexity and a user's mental workload: A NeuroIS perspective," in *Information Systems and Neuro Science*, vol. 16 of *LNISO*, pp. 107–113, Springer, 2016.

[56] R. Buettner, "A user's cognitive workload perspective in negotiation support systems: An eye-tracking experiment," in *PACIS 2016 Proc.*, 2016. 115.

[57] R. Buettner, I. J. Timm, I. F. Scheuermann, C. Koot, and M. Roessle, "Stationarity of a user's pupil size signal as a precondition of pupillary-based mental workload evaluation," in *Information Systems and Neuro Science*, vol. 25 of *LNISO*, Springer, 2017.

[58] R. Buettner, H. Baumgartl, and D. Sauter, "Microsaccades as a Predictor of a User's Level of Concentration," in *Information Systems and Neuroscience*, vol. 29 of *LNISO*, pp. 173–177, Springer, 2018.

[59] S. Sauer, J. Lemke, W. Zinn, R. Buettner, and N. Kohls, "Mindful in a random forest: Assessing the validity of mindfulness items using random forests methods," *Pers. Individ. Differ.*, vol. 81, pp. 117–123, 2015.

[60] S. Sauer, R. Buettner, T. Heidenreich, J. Lemke, C. Berg, and C. Kurz, "Mindful Machine Learning: Using Machine Learning Algorithms to Predict the Practice of Mindfulness," *Eur. J. Psychol. Assess.*, vol. 34, no. 1, pp. 6–13, 2018.

[61] R. Buettner, "A Classification Structure for Automated Negotiations," in *IEEE/WIC/ACM WI-IAT 2006 Proc.*, pp. 523–530, 2006.

[62] R. Buettner and S. Kirn, "Bargaining Power in Electronic Negotiations: A Bilateral Negotiation Mechanism," in *EC-Web '08 Proceedings*, vol. 5183 of *LNCS*, pp. 92–101, 2008.

[63] R. Buettner, "Cooperation in Hunting and Food-sharing: A Two-Player Bio-inspired Trust Model," in *BIONETICS '09 Proc.*, pp. 1–10, 2009.

[64] J. Landes and R. Buettner, "Argumentation-Based Negotiation? Negotiation-Based Argumentation!," in *EC-Web 2012 Proc.*, vol. 123 of *LNBIP*, pp. 149–162, 2012.

[65] R. Buettner, "Asking both the User's Brain and its Owner using Subjective and Objective Psychophysiological NeuroIS Instruments," in *ICIS 2017 Proceedings*, 2017.

[66] R. Buettner, "Getting a job via career-oriented social networking markets: The weakness of too many ties," *Electronic Markets*, vol. 27, no. 4, pp. 371–385, 2017.

[67] R. Buettner, "Predicting user behavior in electronic markets based on personality-mining in large online social networks: A personality-based product recommender framework," *Electronic Markets*, vol. 27, no. 3, pp. 247–265, 2017.

[68] R. Buettner, M. Hirschmiller, K. Schlosser, M. Roessle, M. Fernandes, and I. J. Timm, "High-performance exclusion of schizophrenia using a novel machine learning method on EEG data," in *IEEE Healthcom 2019 Proc.*, IEEE, 2019, in press.

[69] R. Buettner, J. Fuhrmann, and L. Kolb, "Towards high-performance differentiation between Narcolepsy and Idiopathic Hypersomnia in 10 minute EEG recordings using a Novel Machine Learning Approach," in *IEEE Healthcom 2019 Proc.*, IEEE, 2019, in press.

[70] R. Buettner, D. Beil, S. Scholtz, and A. Djemai, "Development of a machine learning based algorithm to accurately detect schizophrenia based on one-minute EEG recordings," in *HICSS-53 Proc.*, IEEE, 2020, in press.

[71] R. Buettner, A. Grimmeisen, and A. Gotschlich, "High-performance Diagnosis of Sleep Disorders: A Novel, Accurate and Fast Machine Learning Approach Using Electroencephalographic Data," in *HICSS-53 Proc.*, IEEE, 2020, in press.

[72] T. Rieg, J. Frick, M. Hitzler, and R. Buettner, "High-performance detection of alcoholism by unfolding the amalgamated EEG spectra using the Random Forests method," in *HICSS-52 Proc.*, pp. 3769–3777, IEEE, 2019.

[73] R. Buettner, T. Rieg, and J. Frick, "Machine Learning based Diagnosis of Diseases Using the Unfolded EEG Spectra: Towards an Intelligent Software Sensor," in *Information Systems and Neuroscience*, vol. 32 of *LNISO*, Springer, 2019, in press.

[74] R. Buettner, T. Rieg, and J. Frick, "High-performance detection of epilepsy in seizure-free EEG recordings: A novel machine learning approach using very specific epileptic EEG sub-bands," in *ICIS 2019 Proc.*, AIS, 2019, in press.

[75] R. Buettner, L. Bachus, L. Konzmann, and S. Prohaska, "Asking Both the User's Heart and Its Owner: Empirical Evidence for Substance Dualism," in *Information Systems and Neuroscience*, vol. 29 of *LNISO*, pp. 251–257, Springer, 2018.

[76] R. Buettner and M. Schunter, "Efficient machine learning based detection of heart disease," in *IEEE Healthcom 2019 Proc.*, IEEE, 2019, in press.

[77] A. Eckhardt, C. Maier, and R. Buettner, "The Influence of Pressure to Perform and Experience on Changing Perceptions and User Performance: A Multi-Method Experimental Analysis," in *ICIS 2012 Proc.*, 2012.

[78] R. Buettner, "Cognitive Workload of Humans Using Artificial Intelligence Systems: Towards Objective Measurement Applying Eye-Tracking Technology," in *KI 2013 Proc.*, vol. 8077 of *LNAI*, pp. 37–48, 2013.

[79] A. Eckhardt, C. Maier, J. P.-A. Hsieh, T. Chuk, A. B. Chan, A. B. Hsiao, and R. Buettner, "Objective measures of IS usage behavior under conditions of experience and pressure using eye fixation data," in *ICIS '13 Proc.*, 2013.

[80] R. Buettner, "Social inclusion in eParticipation and eGovernment solutions: A systematic laboratory-experimental approach using objective psychophysiological measures," in *EGOV/ePart 2013 Proc.*, vol. P-221 of *LNI*, pp. 260–261, GI, 2013.

[81] R. Buettner, B. Daxenberger, A. Eckhardt, and C. Maier, "Cognitive Workload Induced by Information Systems: Introducing an Objective Way of Measuring based on Pupillary Diameter Responses," in *Pre-ICIS HCI/MIS 2013 Proc.*, 2013. Paper 20.

[82] R. Buettner, S. Sauer, C. Maier, and A. Eckhardt, "Towards ex ante Prediction of User Performance: A novel NeuroIS Methodology based on Real-Time Measurement of Mental Effort," in *HICSS-48 Proc.*, pp. 533–542, 2015.

[83] R. Buettner, "Robust user identification based on facial action units unaffected by users' emotions," in *HICSS-51 Proc.*, pp. 265–273, 2018.

[84] R. Buettner, B. Daxenberger, and C. Woesle, "User acceptance in different electronic negotiation systems - a comparative approach," in *ICEBE 2013 Proc.*, pp. 1–8, IEEE, 2013.

[85] R. Buettner, "Towards a New Personal Information Technology Acceptance Model: Conceptualization and Empirical Evidence from a Bring Your Own Device Dataset," in *AMCIS '15 Proc.*, 2015.

[86] R. Buettner, "Analyzing the Problem of Employee Internal Social Network Site Avoidance: Are Users Resistant due to their Privacy Concerns?," in *HICSS-48 Proc.*, pp. 1819–1828, 2015.

[87] R. Buettner, "Getting a Job via Career-oriented Social Networking Sites: The Weakness of Ties," in *HICSS-49 Proc.*, pp. 2156–2165, 2016.

[88] F. Meixner and R. Buettner, "Trust as an Integral Part for Success of Cloud Computing," in *ICIW 2012 Proc.*, pp. 207–214, 2012.

[89] R. Buettner, "The State of the Art in Automated Negotiation Models of the Behavior and Information Perspective," *ITSSA*, vol. 1, no. 4, pp. 351–356, 2006.

[90] R. Buettner, "Electronic Negotiations of the Transactional Costs Perspective," in *IADIS'07 WWW/Internet Proc., Vol. 2*, pp. 99–105, 2007.

[91] R. Buettner, "Imperfect Information in Electronic Negotiations: An Empirical Study," in *IADIS'07 WWW/Internet Proc., Vol. 2*, pp. 116–121, 2007.

[92] J. Landes and R. Buettner, "Job Allocation in a Temporary Employment Agency via Multi-dimensional Price VCG Auctions Using a Multi-agent System," in *MICAI 2011 Proc.*, pp. 182–187, 2011.

[93] R. Buettner and J. Landes, "Web Service-based Applications for Electronic Labor Markets: A Multi-dimensional Price VCG Auction with Individual Utilities," in *ICIW 2012 Proc.*, pp. 168–177, 2012.

[94] R. Buettner, "A Systematic Literature Review of Crowdsourcing Research from a Human Resource Management Perspective," in *HICSS-48 Proc.*, pp. 4609–4618, 2015.

[95] I. Timm, L. Reuter, J. O. Berndt, A.-S. Ulfert, T. Ellwart, and C. Antoni, "Analyzing the Effects of Role Configuration in Logistics Processes using Multiagent-Based Simulation: An Interdisciplinary Approach," in *HICSS-52 Proc.*, pp. 5476–5485, 2019.

[96] B. Zhou, A. Lapedriza, A. Khosla, A. Oliva, and A. Torralba, "Places: A 10 Million Image Database for Scene Recognition," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 40, no. 6, pp. 1452–1464, 2018.