KONFERENZ DER MECHATRONIK PLATTFORM: Autonome mechatronische Systeme
FH CAMPUS 02, 22. November 2018

FH TECHNIKUM WIEN: ARTIFICIAL NEURAL NETWORKS BASED STATE TRANSITION MODELING AND PLACE CATEGORIZATION

Kriegler, A. BSc

Abstract: *To aid the high-level path-planning decisions of a mobile robot, it has to know not just where it is, but also to identify the type and specifics of that place. Customarily, a deep-learning model such as a convolutional neural network (CNN) is trained to classify the location from a video stream. However, a neural network requires fine-tuning and is limited by the closed-set constraint. Following an extensive research review, the aim was to fine-tune an existing system and extend the set of classes it was trained on. The CNN used for feature extraction has been augmented with machine learning (ML) models which have extended the classification and helped to overcome uncertainties from images showing features of multiple classes. The augmented system outperforms the neural network by correctly classifying 90% of images instead of only 78%.*

Keywords: *place categorization, neural networks, machine learning, support vector machines, mobile robotics.*

1. INTRODUCTION

Mobile service robots, that act in complex indoor and outdoor environments alongside humans, need to gain understanding of their surroundings that exceeds basic obstacle-avoidance and autonomous navigation (Sunderhauf et al. 2016). By acquiring knowledge of the environment this way, the robot is better qualified to make appropriate high-level decisions when interacting with a person by modulating its behaviour accordingly.

A CNN is a highly regarded method of extracting information from an image and was used in (Khan et al. 2017) for feature extraction and (Pronobis and Jensfelt 2012) and (Sunderhauf et al. 2016) for place categorization and semantic mapping. The paper by Sunderhauf et al. provided the foundation for this thesis.

At the core of a CNN are the convolutional layers and their parameters, that consist of a set of learnable filters or kernels with a small receptive field that extend through the full depth of the input volume (RGB channels of the input image). During the forward pass the cross-correlation, meaning the dot product between the filter and the input, is computed, resulting in a stack of 2-dimensional activation maps. The architecture of the *Places205* neural network used by Sunderhauf et al. is akin to the popular *AlexNet*.

Following classical machine learning theory (Bishop 2016), three ML models were then trained to enable fine-tuning of the system and allow classification of places not known to the CNN.

2. PROBLEM DESCRIPTION

In the work of Sunderhauf et al. the likelihood values computed by the CNN for the expected places are used as final results and simple top-1 classification amongst them was done. The extraction of the values held by the neurons in the softmax layer that only correspond to the set of expected places leads to an inherent information loss and does not resemble proper fine-tuning. This work instead builds on the results from the CNN to create data sets that are then used to train the ML models. It therefore provides the following contributions:

1. The application of a state-of-the-art CNN for classification is shown.
2. A naive Bayes filter (NBF), a multinomial logistic regression (MLR) model and a support vector machine (SVM) were trained and used for fine-tuning and to extend the set of classes.
3. The system was lastly tested in indoor environments and the “digital factory” of the university.

3. MATERIALS AND METHODS

3.1 CNN for Place Categorization

The CNN-database combination *Places205* published by (Zhou et al. 2014) was specifically trained on a database of 205 different place categories, outperformed the *AlexNet* by roughly 8% top-1 accuracy and still puts up more accurate results than the newer version *Places365*. For creation of the data sets, the complete output-vector $\boldsymbol{y}_{Out} \in \mathbb{R}^{205\times1}$ was continuously extracted from the softmax layer of the CNN to create the design-matrix $\boldsymbol{m}_{205} \in \mathbb{R}^{u\times205}$ and corresponding target vector $\boldsymbol{t}_{205} \in \mathbb{R}^{u\times1}$ respectively, where u is the number of analysed frames. The entire data set was split with 75% of the data forming the training set and 25% making up the test set. The data sets and corresponding target vectors were then used to train the ML models.

3.2 Naive Bayes Filter

NBFs are probabilistic classifiers that are based on the Bayes' rule. In the employed algorithm the densities of each feature variable for each class are first estimated giving a matrix of kernels $\boldsymbol{m}_{kernels} \in \mathbb{R}^{K\times205}$ where K is the number of classes. The posteriors are then modelled using the Bayes' rule. Lastly the model classifier is combined with a decision rule, the *maximum a posteriori* rule. The trained multiclass naive Bayes model used the following optimized parameters:

1. Distribution: Kernel (see (Bishop, 2016), pg. 122)
2. Kernel function type: normal (gaussian distribution)
3. Bandwidth: 0.019527

3.3 Multinomial Logistic Regression

MLR is an extension of the logistic regression for multiclass problems to express the probability as linear combination of independent predictor variables. The Log loss function is convex and can be optimized with iterative optimization schemes such as Newton-Raphson and gradient descent (Bishop 2016). The regularization parameter λ was left at its default value:

1. λ: $1 * 10^{-4}$

3.4 Support Vector Machine

A SVM uses the *kernel trick* to separate and classify data with a high-dimensional hyperplane. The trained SVM was a C-SVM where C determines the trade-off between increasing the margin-size of the soft margin and penalizing data points on the wrong side of the margin. The optimized parameters were:

1. Coding: One vs All
2. BoxConstraint: 75.786
3. KernelScale: 0.7278
4. KernelFunction: Radial Basis Function

4. PRACTICAL REALIZATION

The system was evaluated on 6 observable places on university campus: the digital factory (DF) resembling an industrial setting as the *Places205* class *assembly_line*, a lecture theater as *auditorium*, a *corridor*, a *kitchen* and an *office*. To show the capabilities of the system to extend the limited set, data of an additional place was gathered unknown to the CNN: a *door*. The videos were captured using an inexpensive camera carried around by hand. After analysis of the footage with the CNN, the training- and test-set and target vectors were created as explained in section 3.1

5. RESULTS

Even though the restriction of the number of observable classes helps the CNN-based classifier, it is still less accurate than all developed learning models, as can be seen in table 1.

Classification results				
Environment	**CNN-5**	**NBF**	**MLR**	**SVM**
Digital factory	54%	92%	99%	97%
Lecture theater	62%	81%	88%	89%
Corridor	100%	95%	93%	93%
Door	-	100%	100%	100%
Kitchen	100%	83%	89%	87%
Office	75%	79%	63%	73%
Total average	78.2%	88.27%	88.7%	90.2%

Tab. 1: *Accuracies for the different models:* The bottom row gives the weighted average accuracy.

It should be noted that the results of the CNN cannot so easily be compared to those of the ML models: the models are fine-tuned and the CNN does not know the class door. Nevertheless, the performance of the different systems can be measured with these accuracy calculations. The CNN yields robust classification for places with features that are easily distinguishable from the other 5 classes. As soon as the underlying distribution features peaks in multiple classes, the CNN struggles with classification, only posting a 54% accuracy in the DF. All three ML-models give strong results in the DF and only really struggle to categorize the office. This stems from the fact that some frames from the office recording share similarities with the recorded door and were thus classified as a door - the video recording in the office was suboptimal.

6. CONCLUSION AND OUTLOOK

A straight-forward extension to an accurate and easily implemented system used for place categorization has successfully been developed. The semantic information obtained is an important enabler of more advanced robotics tasks, especially human-robot collaboration. To this end, the system was successfully tested in an industrial and office-esque environment. In future works, the system will be improved and extended upon as follows:

1. The system will be further tested using a mobile robot, gathering both camera and laser scanner data.
2. Using the data from the laser scanner, a semantic map can be created.
3. Path-planning decisions can then be made using this semantic map.

7. BIBLIOGRAPHY

Bishop, M., 2016. *PATTERN RECOGNITION AND MACHINE LEARNING*, Springer-Verlag, 978-0387310732, New York

Khan, A.; Zhang, C.; Atanasov, N.; Karydis, K.; Kumar, V. & Lee, D., 2017. Memory augmented control networks, *Available from: http://arxiv.org/pdf/1709.05706.pdf Accessed: 2018-08-12*

Pronobis, A.; Jensfelt, P., 2012. Large-scale semantic mapping and reasoning with heterogeneous modalities, *Proceedings of the 2012 IEEE International Conference on Robotics and Automation (ICRA)*, pp. 3515-3522, 978-1-4673-1405-3, Saint Paul, MN, USA, 14. – 18. May 2012, IEEE

Sunderhauf, N.; Dayoub, F.; McMahon, S.; Talbot, B.; Schulz, R.; Corke, P.; Wyeth, G.; Upcroft, B. & Milford, M., 2016. Place categorization and semantic mapping on a mobile robot, *Proceedings of the 2016 IEEE International Conference on Robotics and Automation (ICRA),* pp. 5729-5736, 978-1-4673-8026-3, Stockholm, Sweden, 16. – 21. May 2016, IEEE

Zhou, B.; Lapedriza, A.; Xiao, J.; Torralba, A. & Oliva, A., 2014. Learning deep features for scene recognition using places database, *Advances in neural information processing systems,* pp. 487 -495, Montreal, Canada, 08. – 13. December2014,NIPS

KRIEGLER, ANDREAS

Andreas Kriegler BSc

FH Technikum Wien, Höchstädtplatz 6, 1200 Wien, +43-678-1213219, mr18m016@technikum-wien.at

Andreas Kriegler hat im Jahre 2018 mit ausgezeichnetem Erfolg das Bachelor-Diplom im Studiengang Mechatronik/Robotik von der FH Technikum Wien erhalten. Im Zuge der zweiten Bachelorarbeit begann die Forschung an neuronalen Netzwerken für Applikationen in mobiler Robotik, die im Master fortgesetzt wird.