Chapter 7

Conclusion and Further Research

The present work introduces a memory framework for object recognition and categorization systems based on dynamically assembled object models \[94\]. The properties of the proposed computational theory are grounded in the unsupervised structural organization of these object models’ components according to their visual resemblance and co-occurrence, as well as in the use of this structure for matching the novel components. Both have been identified as important for achieving efficient object recognition and categorization and for providing insight on the knowledge-driven aspect of perception \[19\]. These foundations are materialized in three different approaches to the self-organization of visual object knowledge: the Neural Map, the Neural Map Hierarchy (NMH), and the Semantic Correlation Graph (SCG).

Donatti and Würtz \[20\] combine a Growing Neural Gas (GNG) Network \[29, 30\] and a classifier inspired by the coding and decoding of information in the brain \[85\] within a so-called Neural Map. The proficiency of this memory model to self-organize texture information is put to the test by integrating its feature matching responses with a winner-take-all voting scheme. In these experiments, the performance for object recognition and categorization of the resulting Neural Map Classifier (NMC) is validated using image features that derive texture information from object views with different granularity. The present work extends them to borderline Square image features employing different representations to encode gray- and color-valued texture information from object views. It also evaluates alterations on self-organization caused by neural network bootstrapping \[10\]. The overall results indicate that medium-sized image features with the highest amount of feature descriptors maximize the informativeness and distinctiveness of texture information derived from object views. Furthermore, bootstrapping neural network topologies enhances the artificial system’s performance at the cost of slightly longer learning times.

The optimization paradigm introduced in Donatti et al. \[21\] identifies the parameter values of the Neural Map that are best suited for object recognition and categorization. This paradigm
CHAPTER 7. CONCLUSION AND FURTHER RESEARCH

is based on Evolutionary Algorithms [2, 40] (EAs) and explores six different optimization approaches given by the combination of three fitness functions and two starting conditions. The parameter values obtained from the fittest individuals of these approaches are cross-compared and cross-validated with empirical ones using a more elaborate version of the experimental protocol defined in Donatti and Würtz [20]. The parameter values determined with the Sample Distance (SD) fitness function generate Neural Maps that attain the highest object recognition and categorization percentages. The present work also implements bootstrapping and neural growth limits during learning to demonstrate the Neural Map capabilities for data compression.

The search for the fittest individuals encounters two limitations that could be resolved in future research. The first one relates to the computational complexity of training and evaluating large GNG networks using high dimensional image features. This is particularly observed when employing a combination of Global Error (GE) and SD fitness functions with empirical (Emp) starting conditions, which favor individuals that lead to large GNG networks. The evolution of a generation using these processing-intensive approaches can become computationally intractable. The second limitation is the poor ability of the fitness functions to favor the saliency of individuals that may generate better approximations of the feature distribution topology. Donatti et al. [21] introduces the Growth Restriction (GR) to overcome the first limitation, but empirically it is too restrictive, and Neural Maps using this approach are experimentally the least performing ones. As an alternative, the number of generations needed to find an optimal individual can be reduced with a method based on the Covariance Matrix Adaptation (CMA) [39, 82]. However, CMA only optimizes continuous values and, therefore, has to be adapted for discontinuous ones in order to be used with the approach presented in Donatti et al. [21]. Employing CMA might also allow for the use of the CMA for Multi-objective Optimization [45] to restrict the growth rates of the GNG networks. Concerning the second limitation, the optimization process may be improved using additional fitness functions to assess the learning capabilities of a GNG network (e.g., assigning lower fitness values for errors generated in later generations and disregarding errors in earlier ones, or incorporating a rate of change observed in the global error curve of the GNG algorithm).

The NMH is motivated by the efficiency to store and retrieve semantic information of taxonomic hierarchies and by the idea that the cognitive system exploits the tendency of features to occur in clusters across instances in the world. This memory model comprises different abstraction layers with self-organized representations of categories. The representations approximate the topological properties of clusters containing semantically equivalent samples of the feature distribution. The information stored in the more general representations determines the feature weights of more specific levels of categorization. The hierarchy of this memory model is defined according to the taxonomy of the ETH-80 image set [53]. The present work also stud-
ies alternatives employing emergent structures from the unsupervised clustering of the feature distribution using the Enhanced Tree Growing Neural Gas\textsuperscript{72} (ETreeGNG). These experiments overcome the curse of dimensionality\textsuperscript{9} through data quantization and dimensionality reduction techniques, such as Principal Component Analysis\textsuperscript{69} (PCA) and Modified Locally Linear Embedding\textsuperscript{104} (MLLE), which are tailored\textsuperscript{13, 87} to the properties of the feature distribution. Although external validation criteria\textsuperscript{102} and data separability\textsuperscript{22, 50} improve in these embedded distributions, the resulting clustering schemes are not meaningful enough to shape the hierarchy of the NMH. It is worth noting that while using absolute values of feature distances may not be reliable due to the curse of dimensionality, which relates to the challenge of clustering the feature distribution, it is still viable to use rankings of these values for object recognition and categorization\textsuperscript{42}.

Further research in this line of work could apply the MLLE algorithm to non-quantized versions of the feature distribution, provided there are suitable hardware conditions available, or to quantize it with more robust approaches. Additionally, it could also modify the MLLE algorithm utilizing the normalized scalar product instead of the Euclidean distance to define the neighborhoods. Finally, it could also replicate the clustering experiments of the present work with image feature representations that are not based on responses from discrete Gabor filter families, and/or using alternative image libraries having more object categories.

The SCG self-organizes image features depending on their co-occurrence in object views. It employs a configurable weight distribution between occurrences and co-occurrences of image features. The results obtained with different values for this weight distribution are comparable.

The performance of feature-based object recognition and categorization systems equipped with the memory framework introduced in the present work are cross-compared to the ones of state-of-the-art approaches found in literature\textsuperscript{33, 53, 56, 81, 91, 95} utilizing the ETH-80 image set and the COIL-100\textsuperscript{66}. In general, the results for the ETH-80 image set show a gradual decrease of accuracy from the more abstract to the more concrete levels of the taxonomy, reinforcing the concept that the categories from higher abstraction levels are more accurately differentiable, since the characteristics shared by their intra-category individuals are significantly dissimilar in comparison to the ones observed inter-category. In both cases, the artificial systems that employ sequences of object views, either during learning or recall phases, attain the highest object recognition and categorization, in agreement with the intuitive interpretation of visual object understanding in biological systems. An important remark is that the categorization experiments of the present work are obtained through a best case analysis, because novel object views are processed under the same viewing conditions as during training, with near-perfect object segmentation and known scales. Further research should focus on recognition and categorization in more realistic scenes\textsuperscript{26, 41}.