UNIVERSITY OF AMSTERDAM

Nuances in Visual Recognition

E. Gavves

English

The aim of this thesis was to study the fine nuances that could lead to a better visual understanding. Several aspects of this problem have been explored, each of them contributing one more piece to the puzzle. We begin with the question: *is the machine representation of physically identical elements constant?* Our findings support the conclusion that the state-of-the-art image representations using large codebooks can be rather unstable, especially when one attempts to focus on the fine nuances; when going too deep, nuances that correspond to the same physical elements will at some point become too different and in the end will be treated differently. By employing geometry the detected nuances are being described in a stable way, but they are not necessarily optimal as image representations. Hence, we posed our second question: *what set of nuances matters most for object search?* We found that by treating object search from an importance sampling perspective, we are able to discover the few visual nuances that are particularly accurate for certain objects. Encouraged by the result, we investigated the question of *what set of nuances matters most for object classification*. We expressed the problem in terms of regularized least squares. We found that the nuances are not only descriptive among categories, but they are also capable of explaining the contribution of less descriptive nuances. For our next question we addressed \textit{which visual nuances discriminate fine-grained object categories}, where discovering the important nuances is vital and not just helpful for successful classification. Our findings support the idea that fine-grained object categories usually share similar shapes and poses, generally found in their common-super category. Exploiting their common spatial characteristics, fine-grained objects can be examined by their local details from corresponding object regions. We ended up with a highly accurate recognition machine, in a problem considered difficult even for (non-expert) humans. Last, we questioned *whether we can decompose the interpretation of image nuances into local, reusable components, allowing for exact region-level nuance reconstruction*. The results of this thesis show that having the right nuances leads to excellent results. Finding them, however, is non-trivial, it simply depends on what we are looking for.