Abstract

Most convolutional neural networks (CNNs) lack midlevel layers that model semantic parts of objects. This limits CNN-based methods from reaching their full potential in detecting and utilizing small semantic parts in recognition, especially for fine-grained classification.

In this work, we propose a new CNN architecture that integrates semantic part detection and abstraction for fine-grained classification. The detection sub-network has a novel top-down proposal method to generate small semantic part candidates for detection. The classification sub-network introduces novel part layers that extract features from parts detected by the detection sub-network, and combine them for recognition. As a result, the proposed architecture provides an end-to-end network that performs detection, localization of multiple semantic parts, and whole object recognition within one framework. Our method outperforms the state-of-the-art methods with a large margin for small parts detection and also compares favorably to the existing state-of-the-art on fine-grained classification.