

Weakly Supervised Complementary Parts Models for Fine-Grained Image Classification from the Bottom Up

(*Supplementary Material*)

Weifeng Ge^{1,2*}

¹Deepwise AI Lab

Xiangru Lin^{2*}

²The University of Hong Kong

Yizhou Yu^{1†}

1. Ablation Study on Context Fusion.

We perform an ablation study on Stanford Dogs 120 for the context fusion stage. We first replace the multiple losses with the single loss and the accuracy drops from 93.9% to 92.4%. This suggests that multiple losses help regularize the training process and produce more discriminative features for image classification. We then keep the multiple losses setting in subsequent experiments. Second, the Stacked LSTM module is removed and we conduct experiments with two settings, a feature concatenation module and an averaging module. In the feature concatenation module, the features of all the $n + 2$ parts are concatenated. And in the averaging module, the classification output of multiple features are summed. The classification accuracies achieved are decreased by 5.8% and 8.7% respectively. The performance drop suggests that fusing $n + 2$ image patches through LSTM is helpful for final image classification.

2. More Examples of Intermediate Results

We present a list of intermediate results of our pipeline. In the first row of Fig 1, Fig 2 and Fig 3, images from (a) to (e) denote the process of pseudo object mask and object bounding box generation. In the second row, images (f) to (j) indicate the process of Mask RCNN refinement. Fig 4, Fig 5 and Fig 6 denote object parts generated by our proposed complementary parts model.

References

- [1] Gregory Griffin, Alex Holub, and Pietro Perona. Caltech-256 object category dataset. 2007.
- [2] Aditya Khosla, Nityananda Jayadevaprakash, Bangpeng Yao, and Li Fei-Fei. Novel dataset for fine-grained image categorization. In *First Workshop on Fine-Grained Visual Categorization, IEEE Conference on Computer Vision and Pattern Recognition*, Colorado Springs, CO, June 2011.
- [3] P. Welinder, S. Branson, T. Mita, C. Wah, F. Schroff, S. Belongie, and P. Perona. Caltech-UCSD Birds 200. Technical Report CNS-TR-2010-001, California Institute of Technology, 2010.

*These authors have equal contribution.

†Corresponding author is Yizhou Yu.

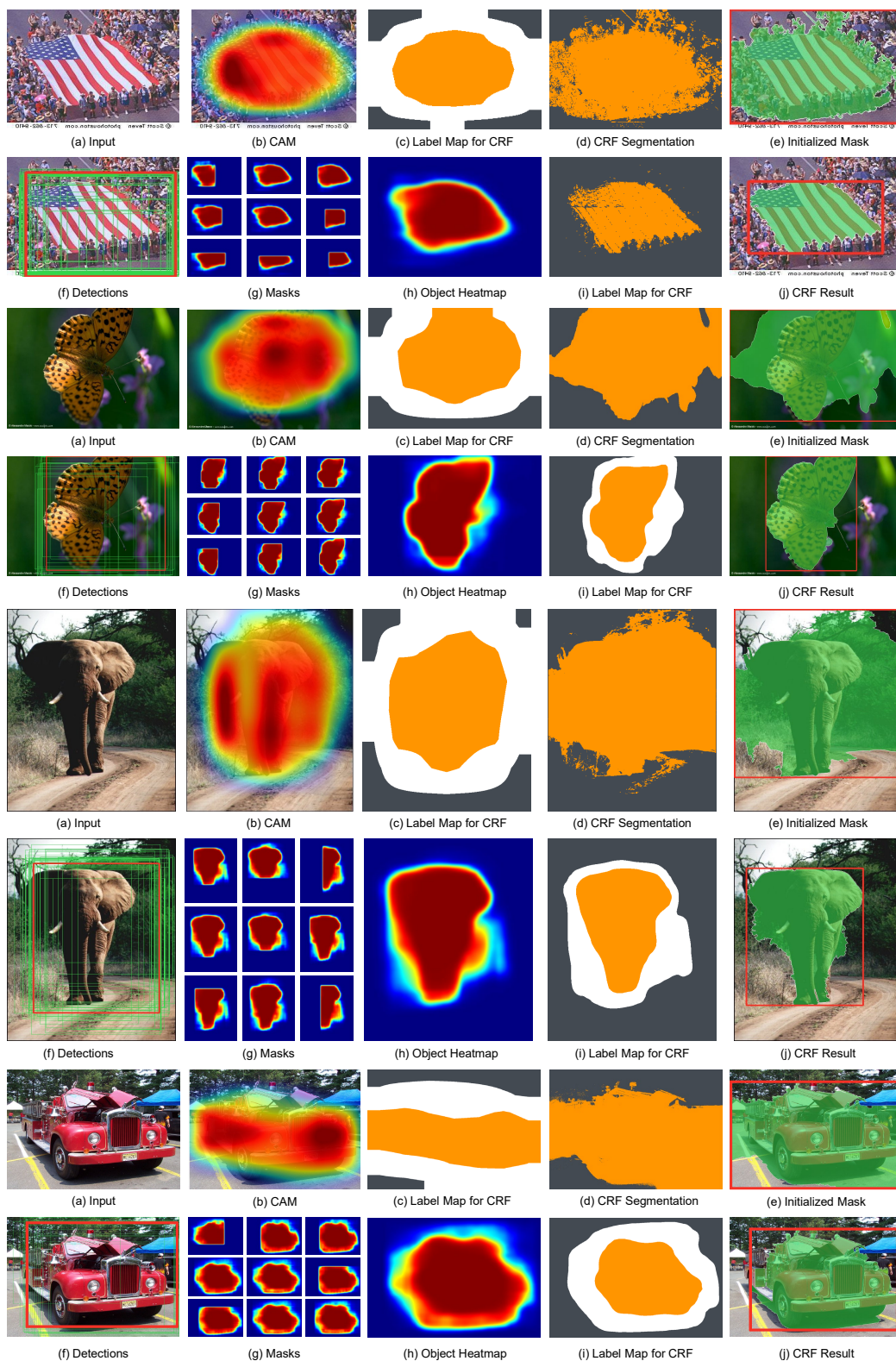


Figure 1. Example intermediate results of Caltech256 [1].

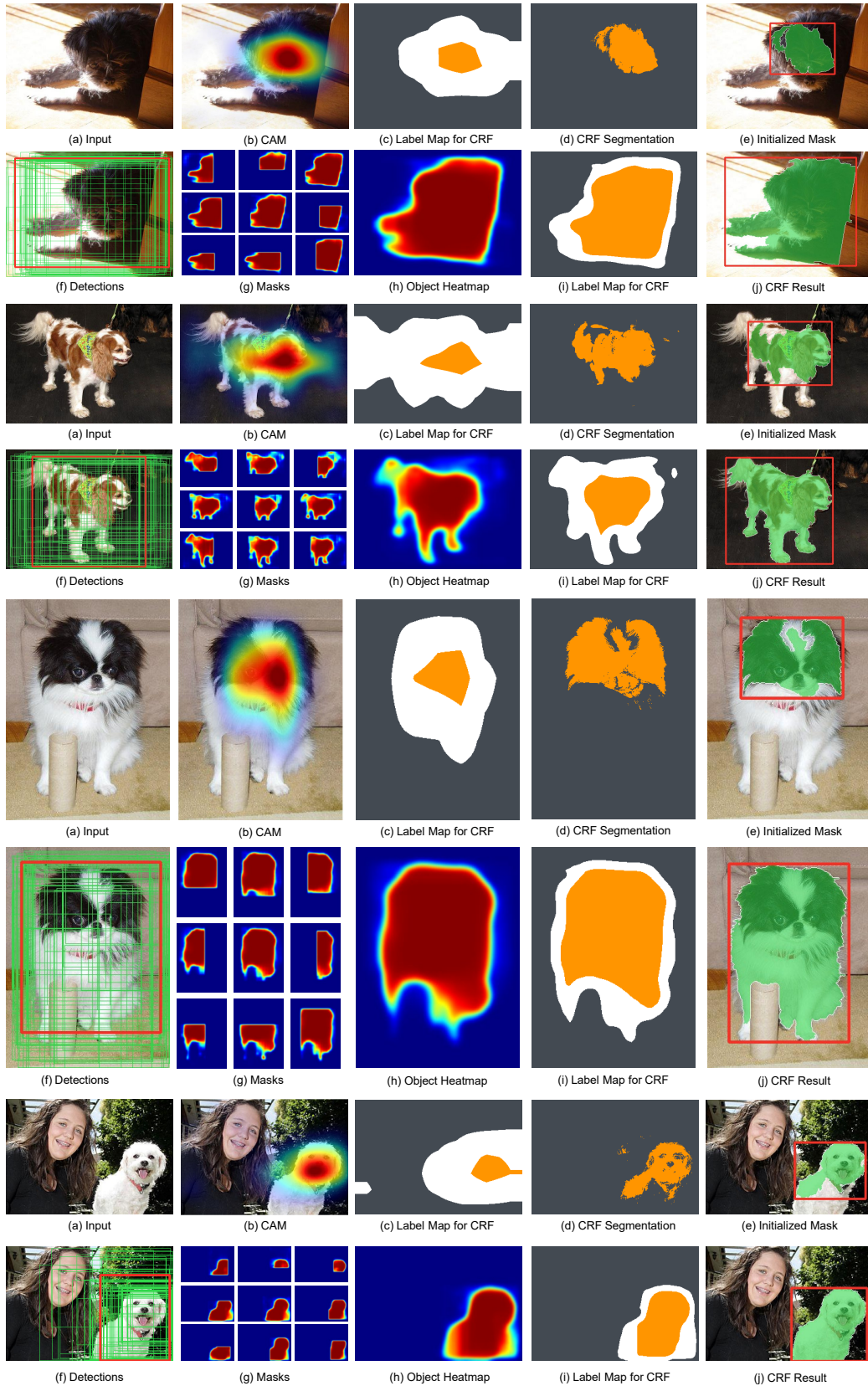


Figure 2. Example intermediate results of Stanford Dogs 120 [2].

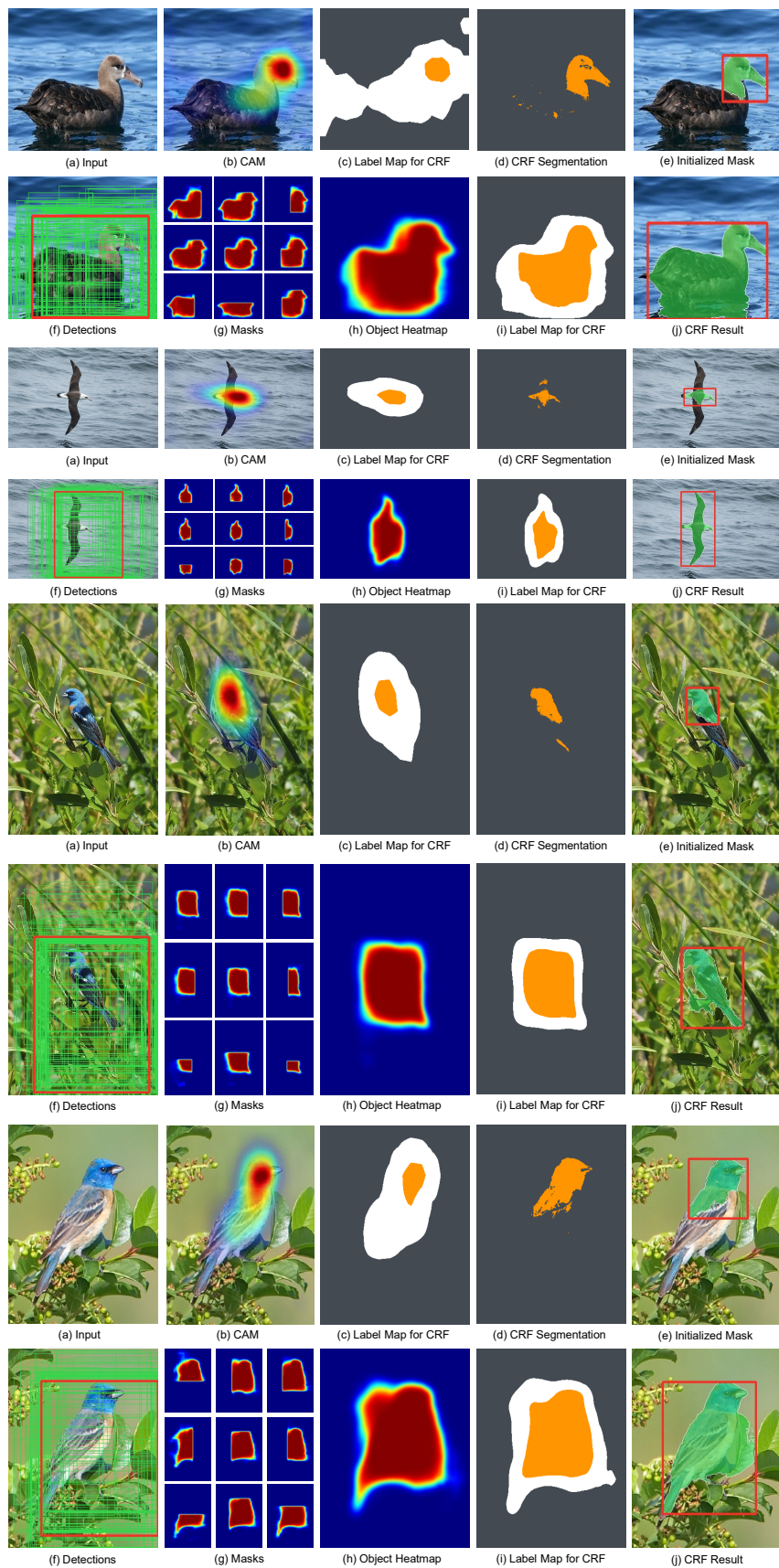


Figure 3. Example intermediate results of CUB 2011-200 [3].

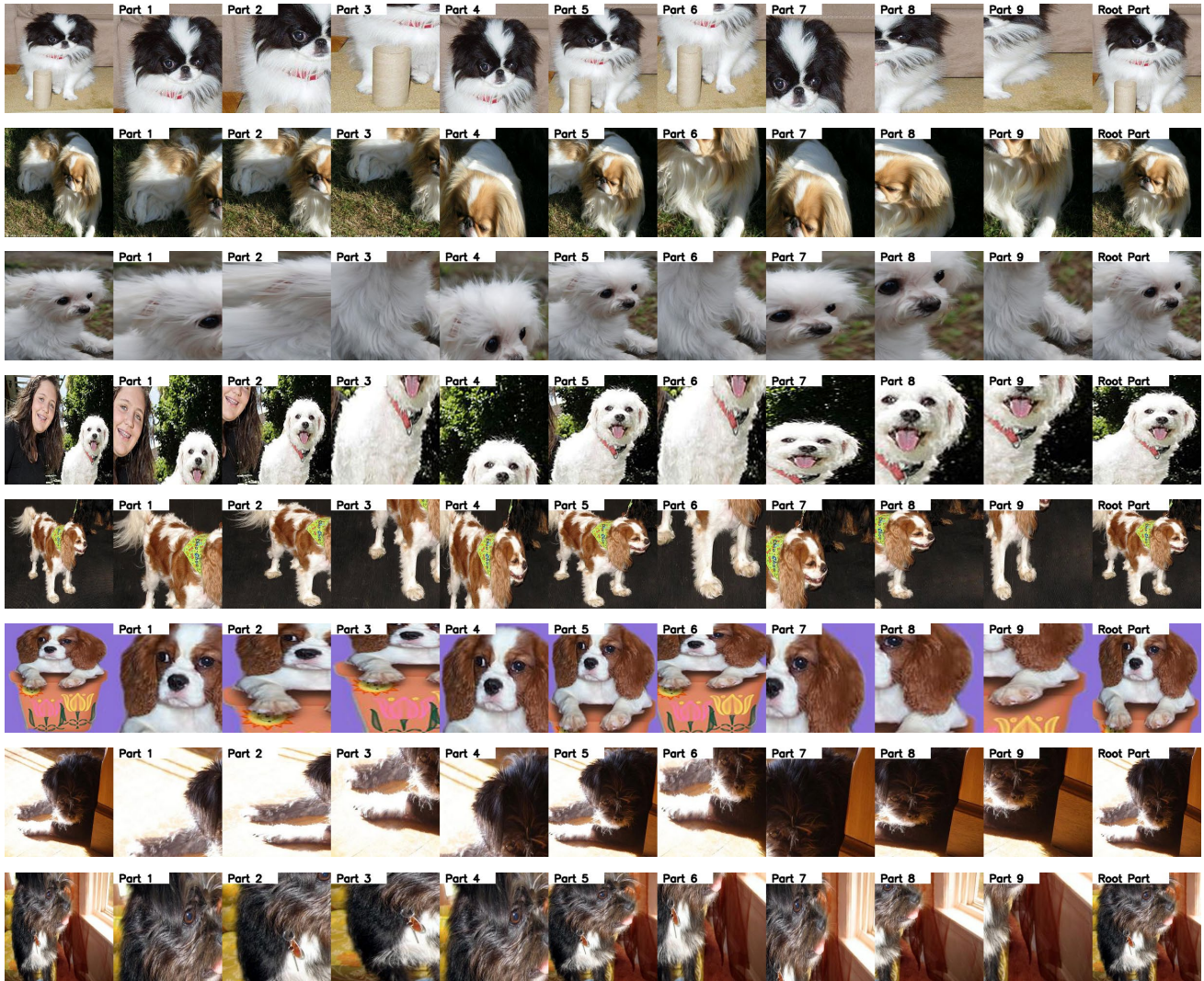


Figure 4. Complementary Parts results of Stanford Dogs 120 [2].

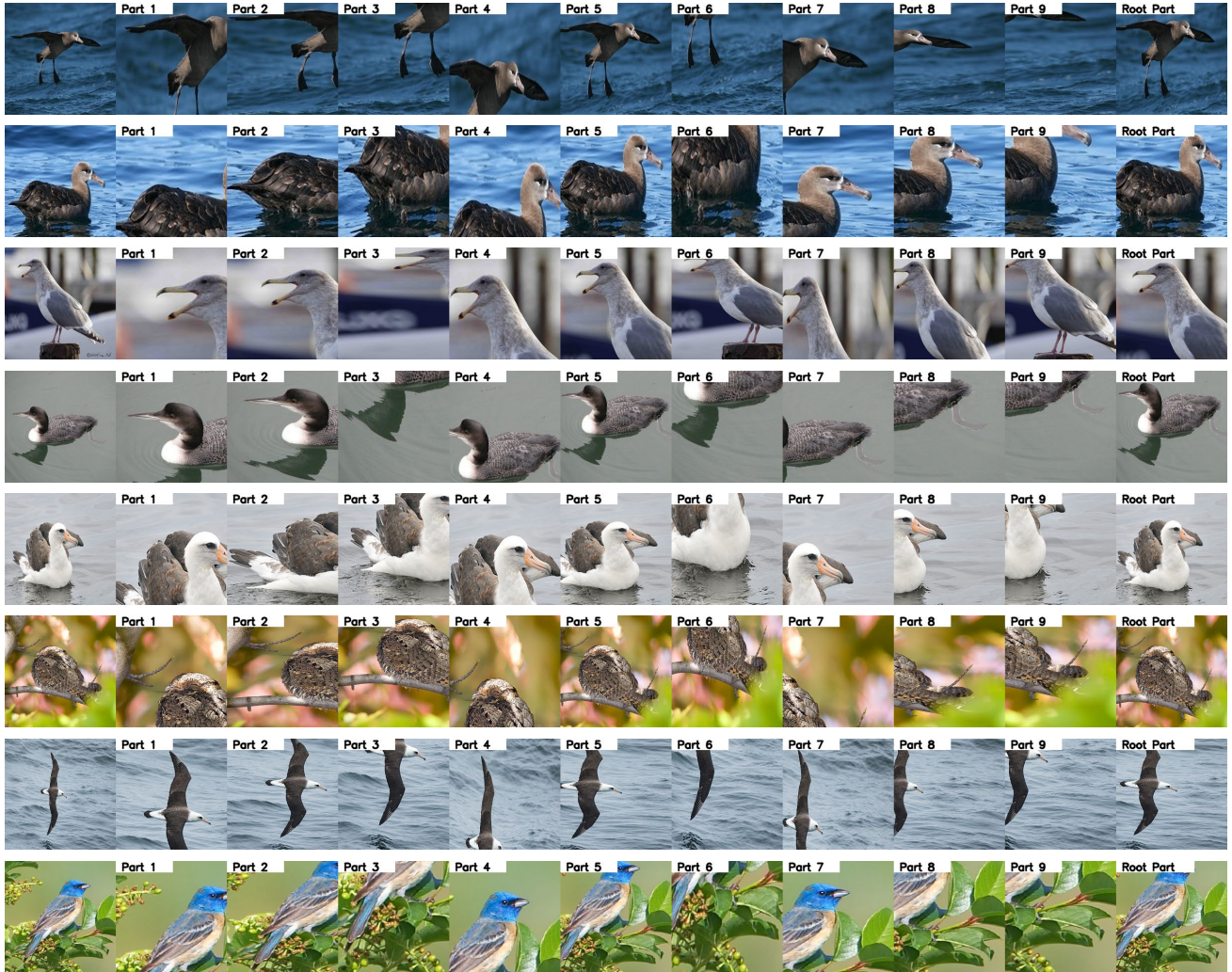


Figure 5. Complementary Parts results of CUB 2011-200 [3].



Figure 6. Complementary Parts results of Caltech256 [1].