High-Level Computer Vision Summer Semester 2021 - Project Proposal Feedback

General Comments

The proposal aims to tackle the fine-grained bird species recognition problem. The motivation is written clearly. The proposal also describes an interesting approach that first detects the bird followed by predicting its species. The dataset and evaluation metrics are reasonable too. I would suggest to develop some naive baselines to better understand where the improvement comes from. Another interesting direction could be zero-shot learning where you train on only 300 bird species and aim to predict the rest of 100 species at the test time [1].

[1] Zero-shot learning - the good, the bad and the ugly. Xian et al., CVPR 2017

Detailed Comments

It would be good to develop the following baselines: (1) the full model without the SPP layer. You could simply crop or resize images to the desired resolution like 224x224. This will essentially tell us how much improvement is attributed to the SPP layer. (2) the full model without the detection network. This will tell us how much improvement is attributed to the detection. You could even develop an oracle experiment to use the ground truth bounding box to see how much we could gain from the perfect localisation. (3) A standard ResNet for this task

Another possible dataset: Caltech Bird. This dataset also provides the attribute annotation for each image. It is also interesting to explore how to use attribute to improve the performance.