ECCV 2018
European Conference on Computer Vision
8 – 14 September 2018 | Munich, Germany

Learning to Navigate for Fine-grained Classification

Ze Yang yangze@pku.edu.cn
Tiange Luo luotg@pku.edu.cn
Dong Wang wangdongcis@pku.edu.cn
Zhiqiang Hu huzq@pku.edu.cn
Jun Gao Jun.gao@pku.edu.cn
Liwei Wang wanglw@cis.pku.edu.cn

Problem

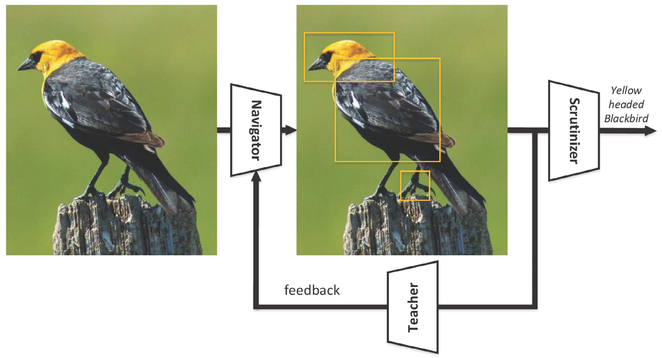
- Fine-grained classification aims at differentiating subordinate classes of a common superior class. The subordinate classes are similar in appearance.


Lazuli Bunting

Indigo Bunting

- The key point to fine-grained classification lies in accurately identifying informative regions in the image.
- Fine-grained human annotations, like annotations for discriminative bird parts, are expensive.
- How can we effectively localize informative regions without the need of fine-grained bounding-box/part annotations?

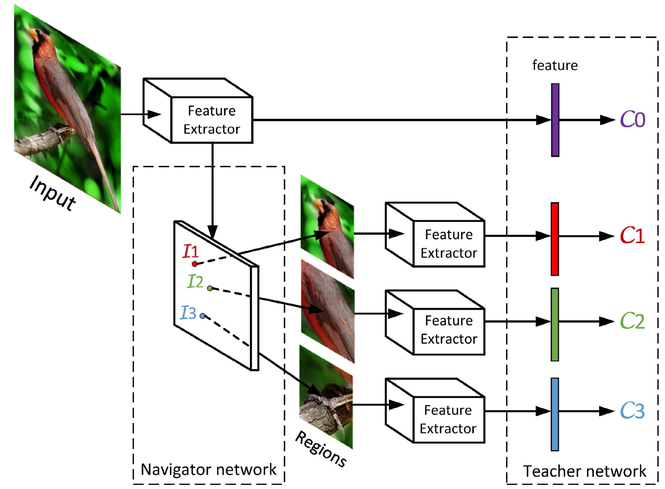
Overview



- Navigator: navigates the model to focus on informative regions.
- Teacher: evaluates the regions and provides feedback.
- Scrutinizer: scrutinizes those regions to make predictions.

Methodology

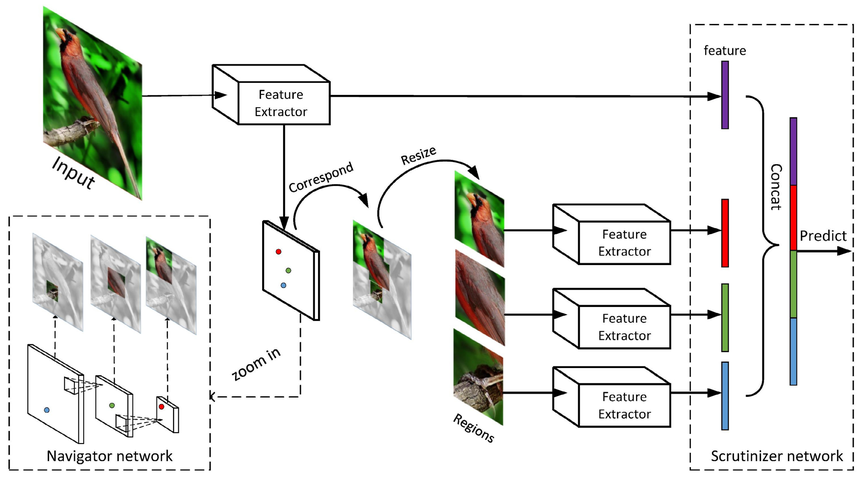
- Train the Navigator to propose informative regions.



Navigator network is a RPN to compute the informativeness of all regions. We choose top-M (here M=3) informative regions with informativeness {I1, I2, I3}. Then the Teacher network compute their confidences being GT class {C1, C2, C3}. We use ranking loss to optimize Navigator network to make {I1, I2, I3} and {C1, C2, C3} having the same order (function *f* is non-decreasing).

$$\text{Ranking loss: } \sum_{(i,s):C_i<C_s} f(I_s - I_i)$$

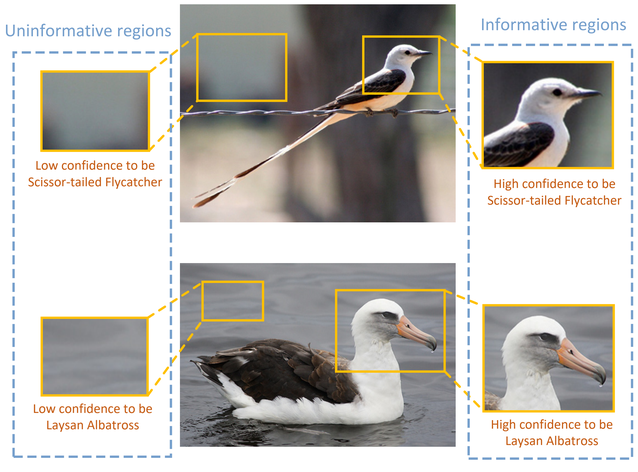
- The Scrutinizer makes predictions.



Navigator network proposes the top-K (here K=3) informative regions. Then the Scrutinizer network uses these regions and full image to make predictions.

Motivations

Intrinsic consistency between informativeness of the regions and their probability being ground-truth class



For informative regions, they will be assigned high probability being GT class. But for uninformative regions that cannot help to differentiate classes, the classifier will not know their class and assigns them low probability being GT class.

Experiments

- Quantitative results.

Method	top-1 accuracy
MG-CNN	81.7%
Bilinear-CNN	84.1%
ST-CNN	84.1%
FCAN	84.3%
ResNet-50	84.5%
PDFR	84.5%
RA-CNN	85.3%
HIHCA	85.3%
Boost-CNN	85.6%
DT-RAM	86.0%
MA-CNN	86.5%
Our NTS-Net (K = 2, M=6)	87.3%
Our NTS-Net (K = 4, M=6)	**87.5%**

Experimental results in CUB-200-2011. The table shows the comparison between our results and previous best results in CUB-200-2011. We use M=6 casually, which means top-6 informative regions are used to train the Navigator. We also study the role of hyper-parameter K, *i.e.* how many part regions have been used for fine-grained classification.

- Qualitative results.

The most informative regions proposed by Navigator network. We can see that the most informative regions are consistent with the human perception

- Birds: head, wings and main body
- Cars: headlamps and grilles
- Airplanes: wings and heads

Especially in the blue box picture where the color of the bird and the background is quite similar.