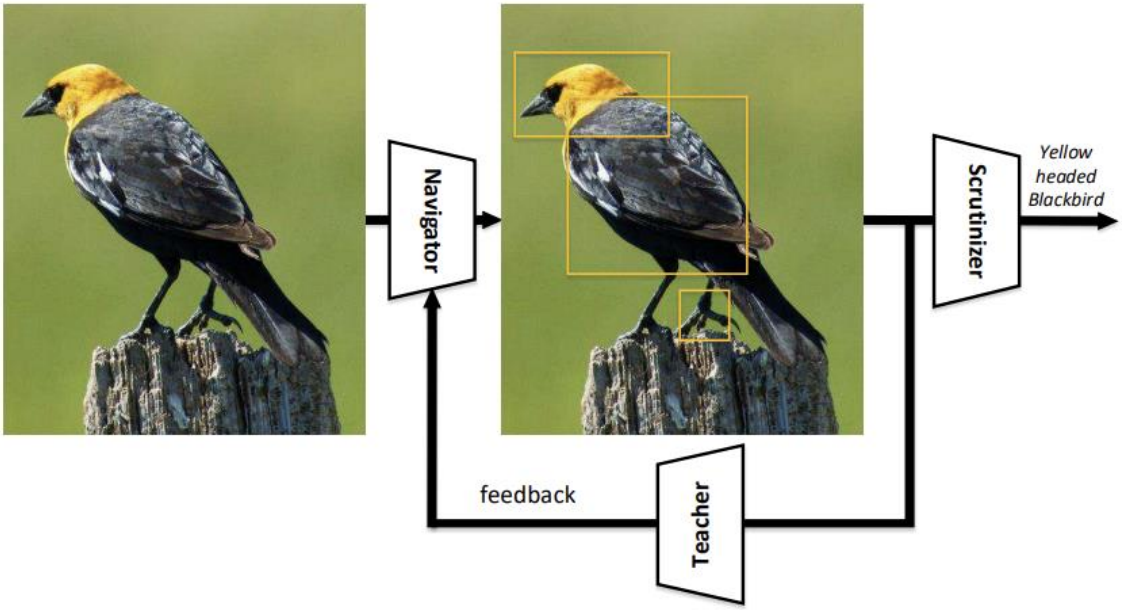
Learning to Navigate for Fine-grained Classification

Bounding-box 필요 없이 정보 영역을 효과적으로 지역화할 수 있는 새로운 메커니즘 제안 NTS-Net(end-to-end로 교육될 수 있으며, 정확한 세분화된 분류 예측, 추론 중 매우 유용한 영역도 제공)

1. Introduction



Navigator는 모델을 탐색하여 가장 유용한 영역(노란색)에 초점을 맞춤. Teacher는 navigator가 제안한 영역을 평가하고 피드백을 제공 그 후, scrutinizer는 예측하기 위해 그 지역을 면밀히 조사

* 이전 방법: human annotation이 필요하기에 비용이 많이 듦 이에 실제 적용 가능성이 낮음.

NTS-Net for Navigator 라는 개발 모델 Teacher-Scrutinizer Network는 이미지에서 정보 영역을 정확하게 식별하는 문제를 해결하기 위해 다중 에이전트 협력 학습 체계사용

ground-truth class이 될 확률이 더 높은 영역은 전체 이미지의 분류 성능을 향상시키는 객체 특성 의미론을 더 많이 포함해야한다. 띠라서 선택된 각 지역의 정보성을 최적화하는 새로운 손실 함수를 설계하여 그 확률이 ground-truth class인 것과 동일한 순서를 갖도록 하고 전체 이미지의 ground-truth class를 영역의 ground-truth class로 함(ground-truth: https://eair.tistory.com/16)

NTS-NET은 Navigator agent, Teacher agent, Scrutinizer agent로 구성된다.

Navigator: 모델의 초점을 가장 유용한 영역에 맞춤. 이미지의 각 영역에 대해 영역이 얼마나 유용한지 예측하고 예측은 가장 유용한 영역 제안

Teacher: Navigaotor가 제안한 지역 평가, 피드백 제공 ground-truth class에 속하는 확률 평가, 신뢰도 평가는 손실함수로 더 많은 정보를 제공할 수 있도록 함(NAvigotor가)

Scrutinizer: Navigator에서 제안된 영역 면밀히 조사

즉, 이 논문에서는

- 바운딩 박스 없이 fine-grained 작업에서 정보 영역을 정확하게 식별하기 위해 새로운 다중 에이전트 협력 학습 체계 제안
- 손실 함수 설계 for Teacher
- End-to-end로 학습될 수 있으며 정확한 fine-grained class 예측, 유용한 영역 제공

2. Related Work

2-1. Fine-grained class

Part annotation없이 바운딩 박스/파트를 행성하기 위해 Co-segmentation(공동 분할)과 alignment(정렬)를 사용하지만 훈련 중에는 바운딩 박스 annotation이 사용됨.

->최근에는 train, inference할 때 바운딩 박스 없어도 되는 방법 등장. 이를 사용할 것임.

네트워크 내의 데이터 표현을 명시적으로 조작하고 정보 영역의 위치를 예측하기 위해 Spatial Transformer Network제안. 전체 이미지의 차별적 특징을 구축하기 위해 bilinearmodel 사용. 모델은 서로 다른 하위 클래스 간의 미묘한 차이를 포착할 수 있다. Bunch of part detectors와 part saliency maps를 학습하기 위해서 "2단계 접근 방식" 제안.[1]

DVAN(Diversified Visual Attention Networ): attention의 다양성을 명시적으로 추구하고 차별적 정보를 더 잘 수집하기 위해

HSNet(Heuristic-Successor Network): 이미지에서 정보 영역에 대한 순처적 검색으로 세분화된 분류 문제를 공식화하기 위해

[1] use an alternate optimization scheme to train attention proposal network and region-based classifier; they show that two tasks are correlated and can benefit each other.

3. Methods

3-1. Approach Overview

정보 영역이 개체의 특성을 보다 잘 나타내는데 도움이 된다는 가정에 의존하기 때문에 정보 영역의 특징과 전체 이미지를 융합하면 더 나은 성능을 얻을 수 있을 것임

- Condition. 1: for any $R_1, R_2 \in \mathbb{A}$, if $\mathcal{C}(R_1) > \mathcal{C}(R_2)$, $\mathcal{I}(R_1) > \mathcal{I}(R_2)$

Navigator network를 자용하여 정보성을 평가하고, Teacher Network를 신뢰기능 C에 적용. Condition.1 을 충족하기 위해 Navigator 최적화함({I(R1),I(R2),...}와 {C(R1),C(R2),...}와 순서 같게 하기 위함)

3-2. Navigator, Teacher

RPN(object detection에서의 핵심역할) 사용

(https://velog.io/@suminwooo/RPNRegion-Proposal-Network-%EC%A0%95%EB%A6%AC)

제안된(?)모든 영역은 groud-truth class와 예측된 신뢰 사이의 cross entropy loss를 최소화하여 teacher를 최적화하는데 사용

3-3. scrutinizer

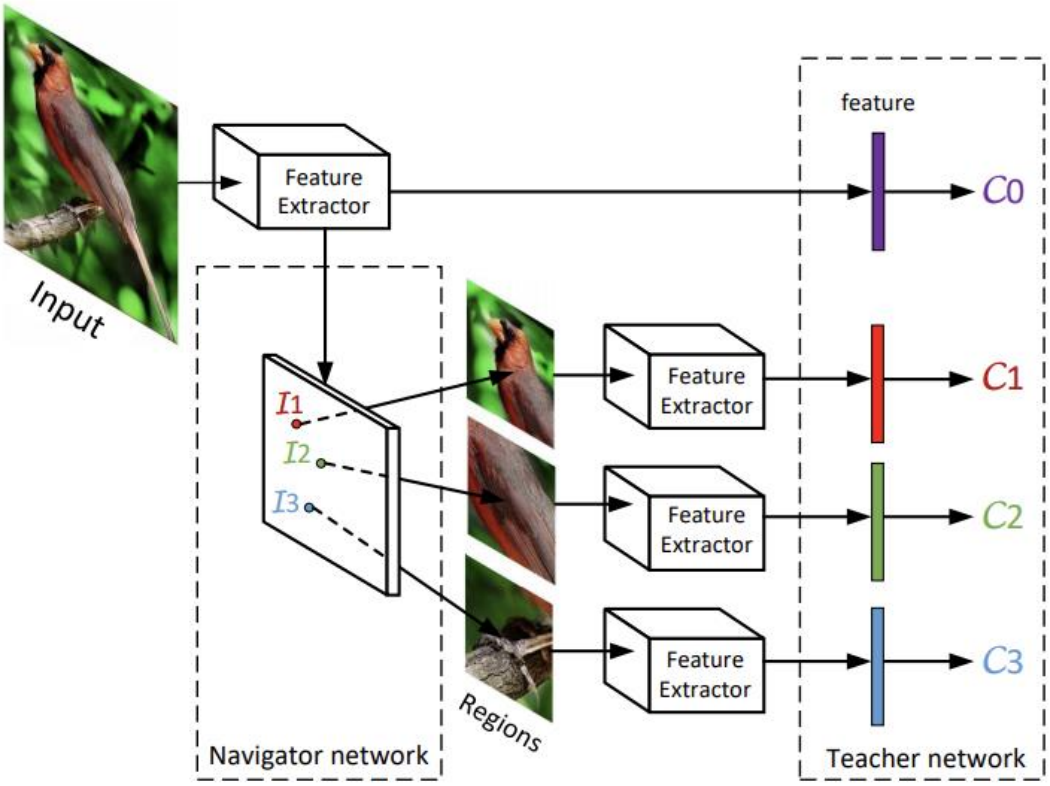
전체 이미지와 결합된 상위 K정보 영역을 입력으로 사용하여 Scrutinizer network 학습. 즉 k 영역은 fine-grained recognition을 용이하게 하기 위해 사용됨. 정보 영역을 사용하여 class 내 분산을 줄이고 올바른 label에 더 놓은 신뢰점수를 생성.

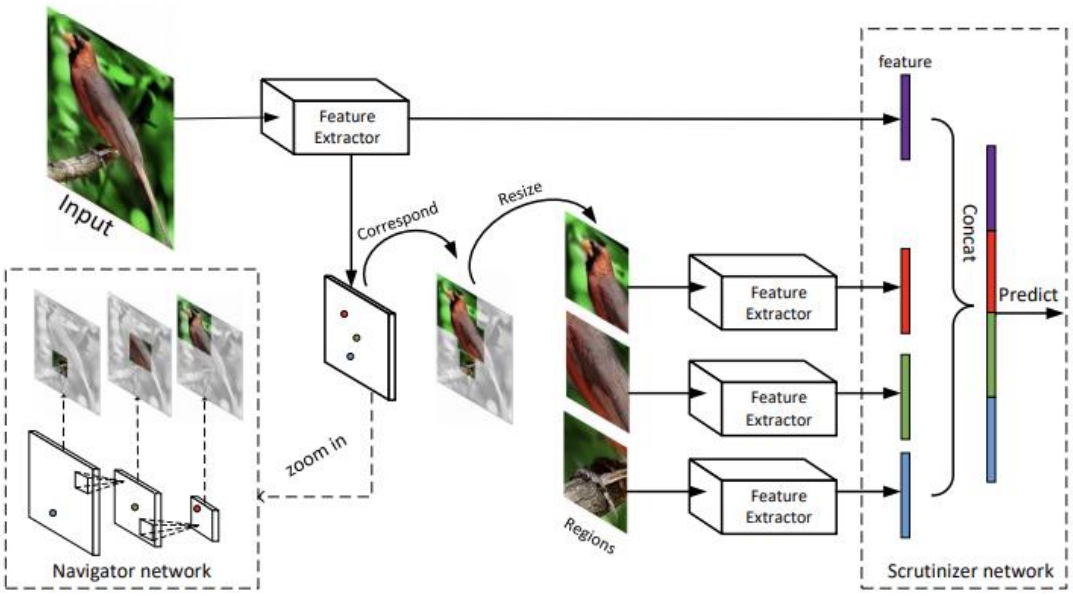
3-4. Network architecture

Navigator

FPN(Feature Pyramid Networks: https://yeomko.tistory.com/44)에서 영감을 받아 다중 스케일 영역을 감지하기 위해 횡방향 연결이 있는 하향식 architecture사용. Convolution 레이어를 사용하여 feature 계층을 계층별로 계산한 다음 ReLU 활성화 함수 및 Max-pooling을 계산.-> 공간 해상도가 다른 일련의 feature map을 얻을 수 있음. 즉 작은 물체, 특징 잡아낼 수 있음.

Teacher





- https://paperswithcode.com/sota/fine-grained-image-classification-on-cub-200