Meta-Learning to Detect Rare Objects

Yu-Xiong Wang Deva Ramanan Martial Hebert

Robotics Institute, Carnegie Mellon University

Objective

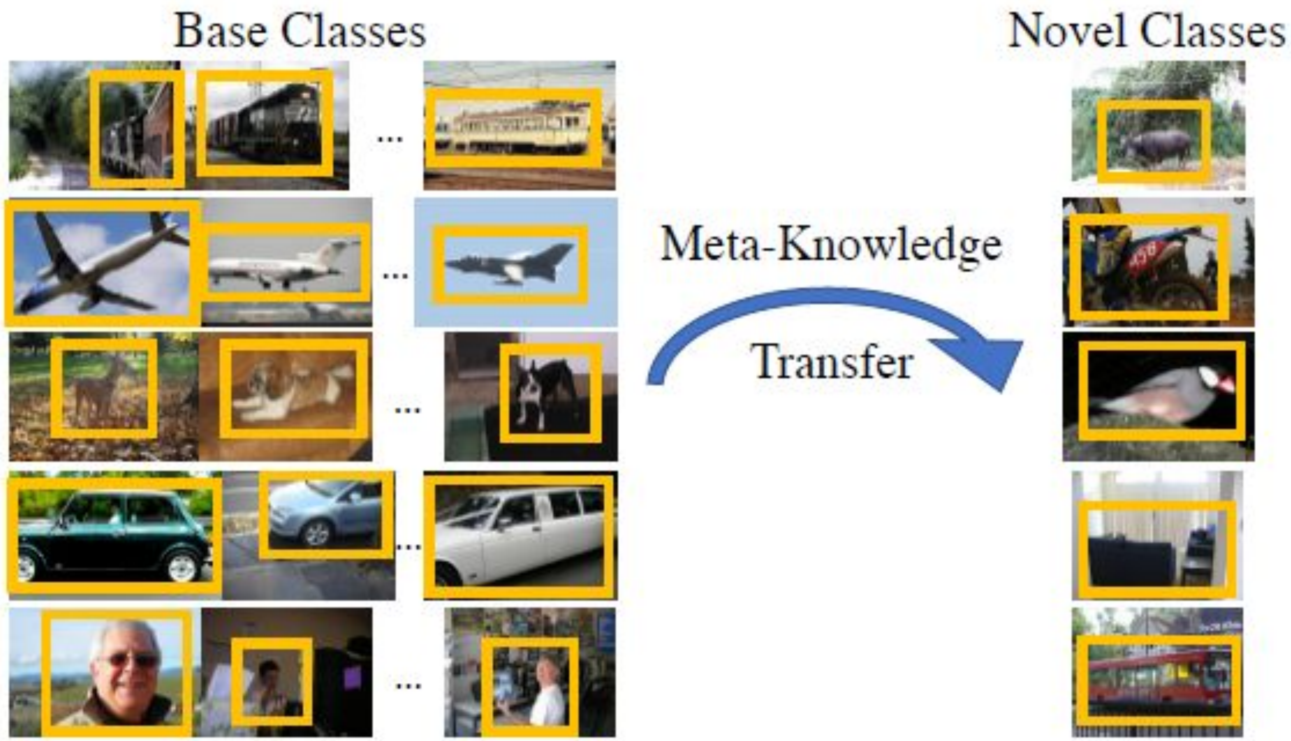
Few-shot learning : Learning novel concepts from few examples

- Tackles few-shot classification and few-shot localization in a unified, coherent way.
- Meta-level knowledge based on model parameter generation from base classes with abundant data to facilitate the generation of a detector for novel classes.

- Disentangle the learning of category-agnostic and category-specific components in a CNN based detection model.
- A weight prediction meta-model that enables predicting the parameters of category-specific components from few examples
- Test the performance of modern detectors in the small-sample size regime.

Few-shot object detection

- Model parameter generation from base classes with a large amount of annotated bounding box examples. (Based on Meta-learning knowledge) The knowledge is then transferred to guide the detector learning for novel classes in a sample-efficient way.



- Disentangle the learning of category-agnostic and category-specific components in a modern CNN based detection model.
- From bottom to top layers of a learned CNN, the model components make a transition from generic to specific.
- Low-level components (bottom convolutional layers), their parameters are shared by many classes and thus category-agnostic. For the high-level components (top fully-connected layers), their parameters tend to be category-specific

Deep Learning based Object Detection

- A few-shot detection model “MetaDet” is built up on Faster CNN.
- Faster R-CNN relies on a region proposal network (RPN) to generate regions of interest (RoIs) on top of convolutional features, and finally uses two sibling branches to classify these RoIs into one of the object classes or background and regress to the refined bounding box positions.

Proposed Method

- Simply transforms a few-shot classifier into a detector, while we simultaneously address few-shot classification and localization.
- It either transfers knowledge from large-sample base set or meta-knowledge from few-shot base detection tasks, but not both.
- Systematically benchmark modern approaches for few-shot detection and significantly outperforms the prior work by large margins.

Proposed Method

- MetaLearning Setup for FewShot Detection: extended the set up for few shot classification to add few shot localization.
- we have a base category set Cbase and a novel category set Cnovel,

$C_{base} \cap C_{novel} = \text{Empty}$

Proposed Method

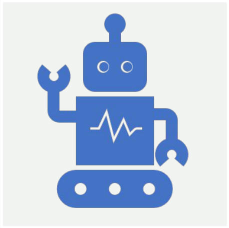
- A large-sample base dataset Sbase = {(Ii, yi)}, where {Ii} are input images, and {yi} are the corresponding annotations indicating labels and bounding boxes for objects of base classes
- k-shot detection, there is a novel dataset Snovel = {(Ii, yi)}, in which each novel class has k bounding box annotations
- The aim is to learn a detection algorithm on Sbase that is able to generalize to unseen categories Cnovel.
- Main focus on the detection performance on Cnovel, which is evaluated on a held-out test set

Proposed method

Through meta-learning, the goal is training a learning procedure (i.e., meta-learner) that guides the generation of a detector (i.e., learner) for a k-shot detection task.

Meta-learning algorithms achieves this by explicitly mimicking the few-shot learning scenario and learning from a collection of k-shot detection tasks sampled from Sbase.

Each of these sampled tasks is termed as *an episode*.



Meta learning algorithms thus have two stages: meta-training on Sbase and meta-testing on Snovel

Proposed Method

During meta-training, there is random sample k bounding box annotations per class on Sbase and train the corresponding detector.

The meta-level knowledge regarding learning detectors across various detection tasks is at the same time aggregated into the meta learner.

During meta-testing, the base detector is adapted on Snovel for novel classes through the meta-learner.

Proposed Method

Basic Detector and Meta Strategies:

- Here we are trying to estimate the parameters ϴ of a detector desired for Cnovel based on both Sbase and Snovel.
- For modern deep CNN based detectors ϴ, are composed of class-agnostic and class-specific component
- To address localization and detection : we use meta-level knowledge about parameter generation of a detection model

Proposed Method

- Here the meta-learning framework applies to Faster R-CNN detectors.
- The detector consists of a region proposal network (RPN) for generating region proposals and a detection network (Fast RCNN) here it uses these proposals to detect objects
- A backbone convolutional network is shared by these two networks and provides convolutional feature maps.

Proposed Method

- RPN identifies region proposals on the feature maps by predicting the probability of an anchor that is the reference box) being foreground or background and refining the anchor.
- These region proposals are then reshaped using a RoI pooling layer, fed into the detection network for predicting the object class by using a softmax classifier and producing the bounding box offsets through the per-class bounding box regressors

Proposed Method

- The convolutional network, RPN, and the bottom layers of the detection network are treated as category-agnostic components, whose parameters are shared by both base and novel classes.
- In the design of Faster R-CNN RPN is category-agnostic, its box-classification layer only assigns a binary class label (of being an object or not) to each anchor without differentiating the specific object classes,
- Also, its box-regression layer regresses from an anchor to a nearby ground-truth box without considering the class label of the ground-truth.
- This shared property enables us to transfer the category-agnostic parameters from the base to the novel detector or use them as initialization for fine-tuning

Proposed Method

- The top layer of the detection network in Faster R- CNN contains category-specific parameters that are used to perform bounding box classification and regression for each class.
- These parameters are not directly transferable between base and novel classes, inspired by the dynamics pattern of how they change from the parameters trained on a small dataset to those trained on a large dataset can be characterized by a eneric, category-agnostic transformation.
- A parametrized weight prediction meta-model T is introduced to learn such a transformation through meta-training process.

Weight Prediction MetaModel

- $w_{\text{det}}^{c,*}$ The class-specific object detection weights in the last layer of the detection network learned from the large-sample base dataset Sbase.
- w_{det}^{c} The corresponding weights learned from the k-shot episode dataset sampled from Sbase
- The weight prediction meta-model T(.) regresses from w_{det}^{c} to $w_{\text{det}}^{c,*}$ in the model parameter space $w_{\text{det}}^{c,*} \approx \mathcal{T}\left(w_{\text{det}}^{c}; \phi\right)$ (where are category-agnostic, learned parameters.

Loss Function

$$||\mathcal{T}(w_{\text{det}}^{c};\phi) - w_{\text{det}}^{c,*}||^2 + \lambda \sum_{(x,y)\in \mathrm{RoI}^c} \text{loss}\Big(\mathcal{D}(x;\mathcal{T}(w_{\text{det}}^{c};\phi)), y\Big).$$

- Final loss is minimized with respect to Φ **(the learned parameter).** which is averaged over all $c \in \mathcal{C}_{\text{base}}$ and over w_{det}^{c} generated in all the episodes.
- Here ‘loss’ refers to the standard performance loss used to train the detection network D.

Example : multi-task loss consisting of bounding box classification and regression

losses), and RoIc denotes the training RoIs labeled with class c. (Region of interest)

Loss Function

- The bounding box detection branch contains two types of detection weights: the RoI classification weights wc cls and the bounding box regression weights wc loc.
- The concatenation of the two types of weights is used, wc dec = [wc cls,wc loc].
- Therefore, we simultaneously address few-shot classification and few-shot localization in a unified way, extending the sole classification
- T (·) can be implemented as a small fully-connected neural network and jointly trained with the detector

Meta learning Procedure

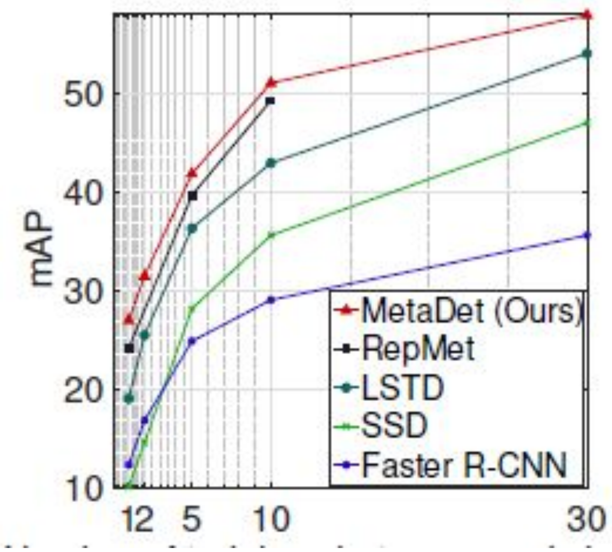
- Consists two phases: meta-training on Strain and meta-testing on Snovel.
- Stage-wise meta-training. Splits the meta-training procedure into two stages for category-agnostic and category-specific components,
- In the first stage, we train a large-sample base detector D(Ɵ) on the entire dataset Strain in the normal way.
- This provides the basic detector which will be used for novel classes and the large-sample parameters of the class specific components.

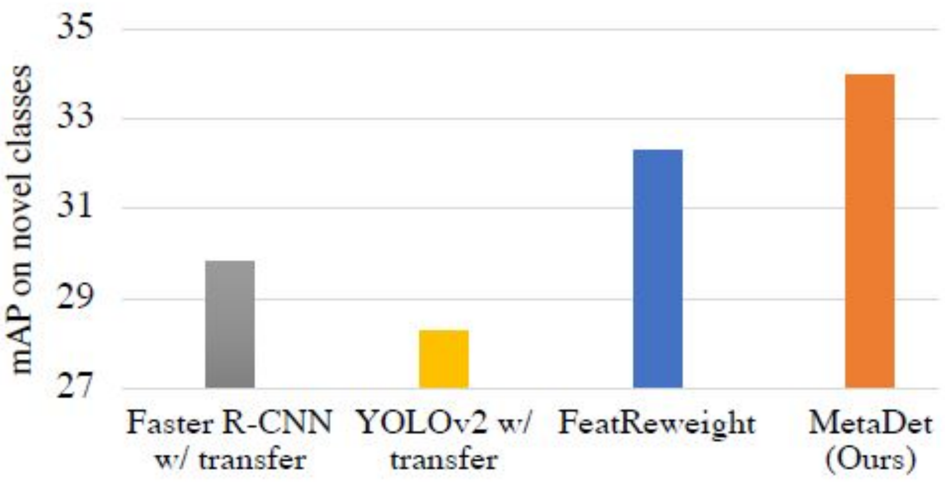
Meta learning Procedure

- On second stage, few-shot episode detection is performed.
- In each episode, we randomly sample k bounding box annotations per class on Strain.
- We leverage the large-sample base detector trained in the first stage to generate the k-shot detector,
- We freeze the category-agnostic parameters as those learned in the large-sample setting and retrain the category-specific parameters from scratch.
- We use k-shot detector together with the k-shot examples to train the meta-model based on the meta-objective and Everything is trained end-to-end

Meta-learning based approach "MetaDet" for few-shot detection tasks was tested on the PASCAL VOC dataset, MS-COCO, ImageNet, and iNaturalist.

Experimental Evaluation





Conclusion

- Fewshot detection for novel classes that simultaneously tackles
- Few-shot classification and localization in a unified, by using meta-learning techniques.
- A strategy was proposed to disentangle the learning of category-agnostic and category-specific components in a CNN based detection model.
- The approach achieves state-of-the-art detection performance in a variety of realistic scenarios, including within-domain, cross-domain, and long-tailed settings, and under different notions of novel classes.

Thank you ☺