Prototypical Faster R-CNN for Few-Shot Object Detection on Aerial Images

Open Session DeepLearn Summer 2021

Pierre Le Jeune, Anissa Mokraoui, Mustappha Lebbah, Hanene Azzag

1. Problem statement

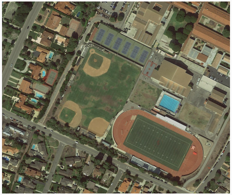
n-way k-shot object detection

Given support examples $\{(x_1, a_1), \ldots, (x_{nk}, a_{nk})\}$ it consists in detecting all occurences of classes in $\mathcal{C}$ ($|\mathcal{C}| = n$) in a query image x_q.

1. Problem statement

n-way k-shot object detection

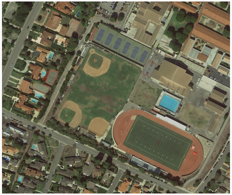
Given support examples $\{(x_1, a_1), \ldots, (x_{nk}, a_{nk})\}$ it consists in detecting all occurences of classes in $\mathcal{C}$ ($|\mathcal{C}| = n$) in a query image x_q.



Query image

n-way k-shot object detection

Given support examples $\{(x_1, a_1), \dots, (x_{nk}, a_{nk})\}$ it consists in detecting all occurences of classes in $\mathcal{C}$ ($|\mathcal{C}| = n$) in a query image x_q.



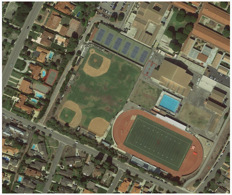
Query image

Support examples

1. Problem statement

n-way k-shot object detection

Given support examples $\{(x_1, a_1), \dots, (x_{nk}, a_{nk})\}$ it consists in detecting all occurences of classes in $\mathcal{C}$ ($|\mathcal{C}| = n$) in a query image x_q.



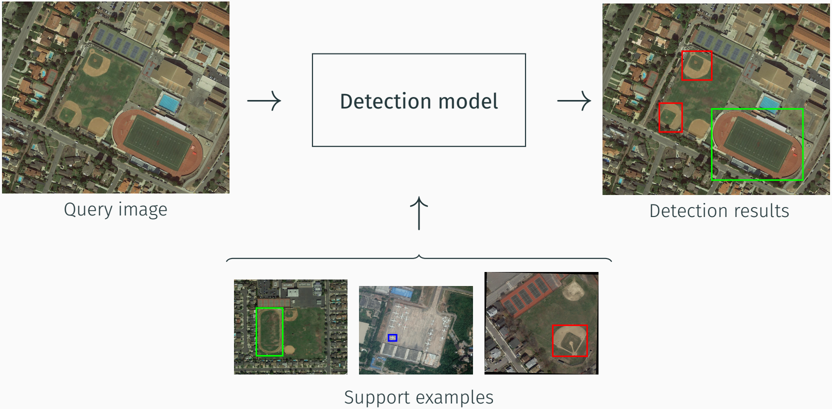
Query image

Detection model

Support examples

n-way k-shot object detection

Given support examples $\{(x_1, a_1), \dots, (x_{nk}, a_{nk})\}$ it consists in detecting all occurences of classes in $\mathcal{C}$ ($|\mathcal{C}| = n$) in a query image x_q.



Our work is astride two research areas: object detection and few-shot learning

Our work is astride two research areas: object detection and few-shot learning

- **Object detection:** Faster R-CNN (Ren et al. 2015)

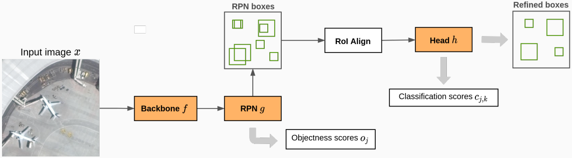


Figure 1: Faster R-CNN.

Our work is astride two research areas: object detection and few-shot learning

- **Object detection:** Faster R-CNN (Ren et al. 2015)

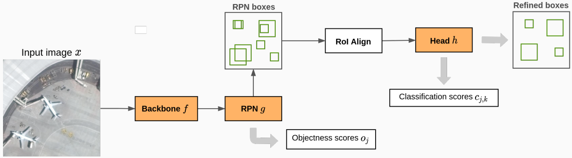


Figure 1: Faster R-CNN.

- **Few-shot Learning:** Prototypical networks (Snell, Swersky, and Zemel 2017)

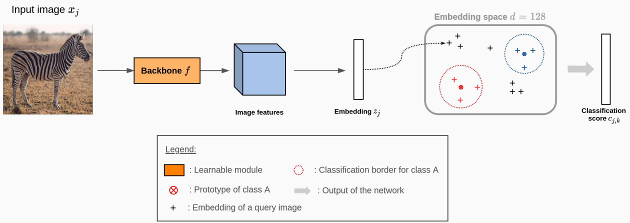


Figure 2: Prototypical networks.

Principle: replace classification branches in Faster R-CNN by prototypical networks.

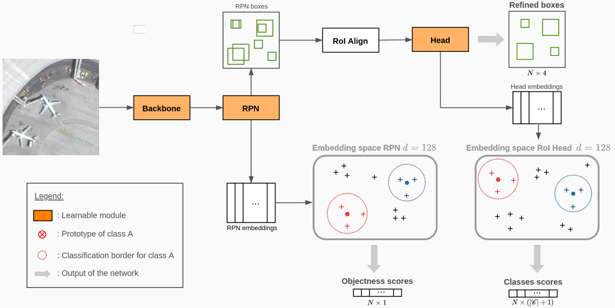


Figure 3: Prototypical Faster R-CNN architecture.

For each generated box b an embedding vector is computed, objectness and classification scores are attributed according to the distances between embeddings and class prototypes.

$$p(x_{j,b} | y_{j,b} = c) = \exp\left(\frac{-d(z_{j,b}, p_c)^2}{2\sigma^2}\right), \qquad (1)$$

$$o_{j,a} = \max_{c \in C_j} p(x_{j,b} | c), \qquad (2)$$

$$p(c | x_{j,b}) = \frac{p(x_{j,b} | c)}{\sum\limits_{c \in C_j \cup \{\varnothing\}} p(x_{j,b} | c)}. \qquad (3)$$

Experimental protocol: some classes are retained from the dataset during training (test classes). At test time performance is assessed both on training and test classes.

	k shots	1	3	5	10
Split A	Train classes	0.275 ± 0.01	0.352 ± 0.02	0.390 ± 0.01	0.384 ± 0.02
	Test classes	0.047 ± 0.02	0.024 ± 0.01	0.038 ± 0.01	0.041 ± 0.01
Split B	Train classes	0.415 ± 0.03	0.392 ± 0.03	0.434 ± 0.02	0.414 ± 0.03
	Test classes	0.08 ± 0.01	0.101 ± 0.02	0.121 ± 0.01	0.101 ± 0.02

Table 1: Mean average precision over 5 runs on DOTA dataset with 95% confidence interval. Results are given for two different train/test classes split. Split A: $\{0, 1, 4\}$, Split B: $\{7, 11, 13\}$ (only test classes are given).

- Low performance, especially for test classes explained by overlapping cluster in representation space.
- For training classes performance slightly lower than regular training. Clusters well separated.

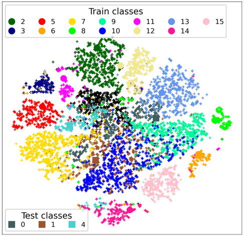


Figure 4: TSNE projection of the embeddings

Key difficulties and potential solutions:

5. Perspectives

Key difficulties and potential solutions:

- Faster R-CNN is not well suited for Few-shot object detection
  - Low performance in RPN can cripple classification head.
  - Two-stages and anchors approach bring unecessary complexity.

Key difficulties and potential solutions:

- Faster R-CNN is not well suited for Few-shot object detection
  - Low performance in RPN can cripple classification head.
  - Two-stages and anchors approach bring unecessary complexity.

  $\rightarrow$ Change detection framework with one stage end-to-end, e.g. FCOS (Tian et al. 2019).

Key difficulties and potential solutions:

- Faster R-CNN is not well suited for Few-shot object detection
  - Low performance in RPN can cripple classification head.
  - Two-stages and anchors approach bring unecessary complexity.
  - → Change detection framework with one stage end-to-end, e.g. FCOS (Tian et al. 2019).
- Support examples are unlikely to match with query image
  - Semantic information may be dominated by background within a patch.
  - Classification scores depend on background similarity between example and patch.

5. Perspectives

Key difficulties and potential solutions:

- Faster R-CNN is not well suited for Few-shot object detection
  - Low performance in RPN can cripple classification head.
  - Two-stages and anchors approach bring unecessary complexity.

  $\rightarrow$ Change detection framework with one stage end-to-end, e.g. FCOS (Tian et al. 2019).
- Support examples are unlikely to match with query image
  - Semantic information may be dominated by background within a patch.
  - Classification scores depend on background similarity between example and patch.

  $\rightarrow$ Adapt prototypes to match embeddings through attention mechanism.

Thank you for your attention

Any questions

pierre.lejeune@lipn.univ-paris13.fr

Ren, Shaoqing et al. (2015). "Faster r-cnn: Towards real-time object detection with region proposal networks". In: *Advances in neural information processing systems* 28, pp. 91–99.

Snell, Jake, Kevin Swersky, and Richard Zemel (2017). "Prototypical Networks for Few-shot Learning". In: *Advances in Neural Information Processing Systems*. Ed. by I. Guyon et al. Vol. 30. Curran Associates, Inc. URL: https://proceedings.neurips.cc/paper/2017/file/cb8da6767461f2812ae4290eac7cbc42-Paper.pdf.

Tian, Zhi et al. (2019). "Fcos: Fully convolutional one-stage object detection". In: *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pp. 9627–9636.