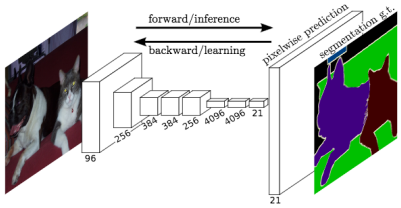
Model Card for LCFCN

Intended use

LCFCN is a deep learning model that can count and localize objects of interest for a given image. Its goal is to output a blob on each object that it counts. Its applications range from animal conservation (demo) to crowd security measures.



Model architecture of FCN8.

Output of LCFCN fine-tuned on Penguins

Model architecture

The model could use any architecture that can output a per-pixel prediction for a given input image. In this work, the model uses a Fully Convolutional Network (FCN8) with skip connections and a loss function that has four loss terms: image-level, point-level, split-level, and false positive loss, which collectively encourage the model to output a single blob per object during training.

Metrics

Object Counting: mean absolute error and (weighted) root mean squared error.
Localization: Grid Average Mean absolute Error (GAME) and F1-score where true positive blobs must cover at least 1 dot annotation.
Lower is better for all error metrics while higher is better for F1-score.

Model Performances

We report the mean absolute error. See §4 of the paper for the other metrics and the other datasets. Additional results on synthetic data can be found in this paper.

	MIT Traffic	PKLot	Trancos	Penguins
Glance	1.57	1.92	7.01	6.09
LCFCN	**0.91**	**0.21**	**4.53**	**3.74**

Factors and limitations

Images that are small (less than 128 x 128) and have highly overlapping objects could lead to a decrease in performance. Further, testing the trained model on a camera not used in training could increase the MAE by around twice as much.

Unusual orientation led to multiple blobs

Penguins in the top third are too small

Ethical considerations

People did not consent to their image being used in ImageNet or the crowd datasets. ImageNet's distribution of people and biodiversity is biased, thus fine-tuned versions of LCFCN may carry those biases. Other harms include surveillance and invasion of privacy. Military usage of this model is possible as well.

Model details

Released: July 2018
Resources: **Paper** **Repo** **BibTeX**
License: Apache 2.0
Contact: Queries can be addressed on our **repo**.

Evaluation data

Penguins contains hourly images of penguins taken by fixed cameras in over 40 locations in Antarctica.

Traffic datasets: **Trancos** has images of highly overlapping vehicles from multiple roads and **MIT Traffic** contains videos from a single camera.

PKLot contains surveillance images of parking spaces.

Pascal VOC 2007 is a standard dataset for object class recognition.

Crowd datasets: **UCSD** contains videos from a fixed camera overlooking a pedestrian walkway, **Mall** contains video frames of pedestrians with diverse gender and ethnicity in shopping malls, and **ShanghaiTech B** contains high-resolution images.

Training data

The model is pre-trained on **ImageNet** and fine-tuned on the above datasets.

Data pre-processing

Normalization with ImageNet mean and standard deviation.

Additional information

You can try LCFCN in **Github** and **Colab**.