



Deep Image Prior; DOCK

Presented by Divya K Raman

Computer Vision Lab, IIT Madras

25 Jan 2019



Schedule

- ❖ Paper 1: Detecting Objects by Transferring Common Sense Knowledge
- ❖ Paper 2: Deep Image Prior



Paper 1

DOCK: Detecting Objects by Transferring Common sense knowledge

Author(s): KK Singh et al; Published at ECCV 2018

Additional References: <https://dock-project.github.io/>



Introduction


- Approach for Detecting Objects by Transferring Common sense knowledge(DOCK) from source to target categories
- Training data: source categories – bounding box annotations, target categories – image level annotations
- Key ideas: (i) use similarity at region level (ii) leverage richer common-sense

Challenges Addressed

- Previous transfer learning methods - similarity knowledge between the source and target categories, insufficient.



Answer: Toothbrush

- 
- Previous methods – need robust image level object classifier for transferring knowledge
 - if instances of the target classes frequently co-occur with the source classes, then the target class regions can end up being undesirably learned as ‘background’ while training the detector for the source classes.
 - Overcoming first limitation: leverage multiple sources of common-sense knowledge. Specifically, they encode: (1) similarity, (2) spatial, (3) attribute, and (4) scene.
 - Addressing limitation 2 and 3: directly model objects at the region-level rather than at the image-level.



Approach

Part 1: Base Detection Network

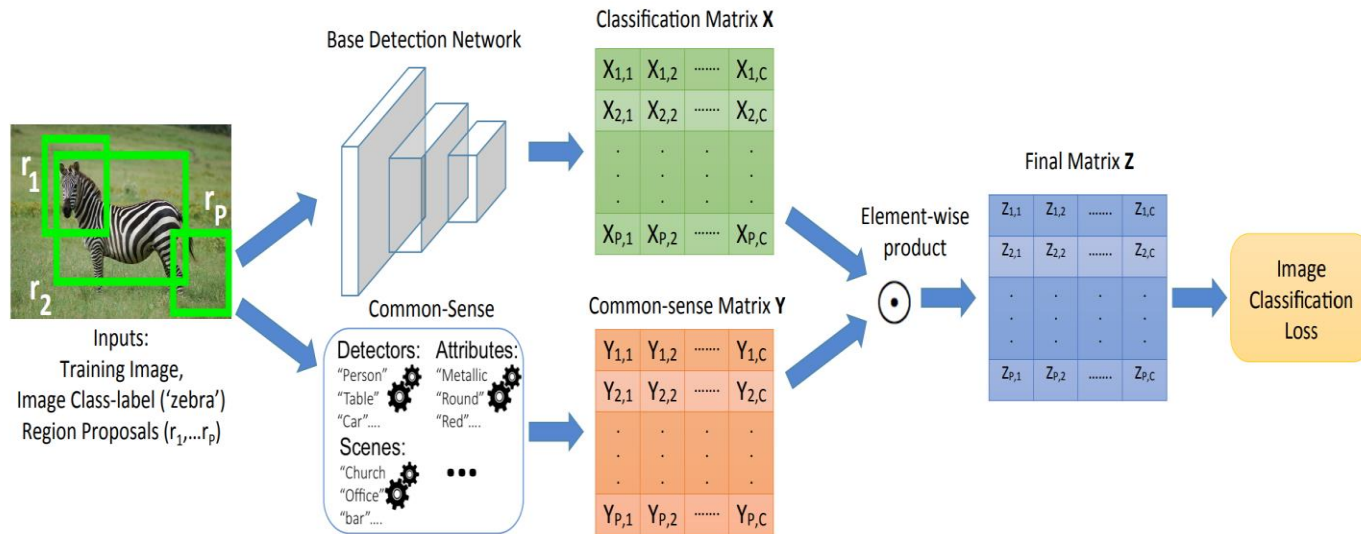
- Initial layers of the network – convolution layers
- Followed by spatial pyramid pooling layers to pool features corresponding to image region proposals
- After pooling, the network has two data streams
- Recognition stream: assigns a classification score for each region proposal by applying a softmax over the classes to produce a $P \times C$ recognition matrix
- Detection stream: assigns probability of a region proposal to be selected for a specific class by applying a softmax over the proposals to produce a $P \times C$ detection matrix
- The final probability for each proposal to belong to different classes is computed by taking element-wise dot product of detection and recognition matrix
- The network takes P proposals of a training image as input and outputs the probability for each of them to belong to C classes.
- The image-level class probabilities are obtained by summing the probabilities of each class over the proposals



Approach

Part 2: Transferring common sense

- Augment the above base detection network with a common-sense matrix Y of size $P \times C$
- Each element of Y : represents a 'prior' probability of a proposal belonging to a class according to common-sense knowledge
- Separate common-sense matrix for each type of common-sense (similarity, attribute)
- Element wise dot product of common sense matrix with classification matrix from part 1

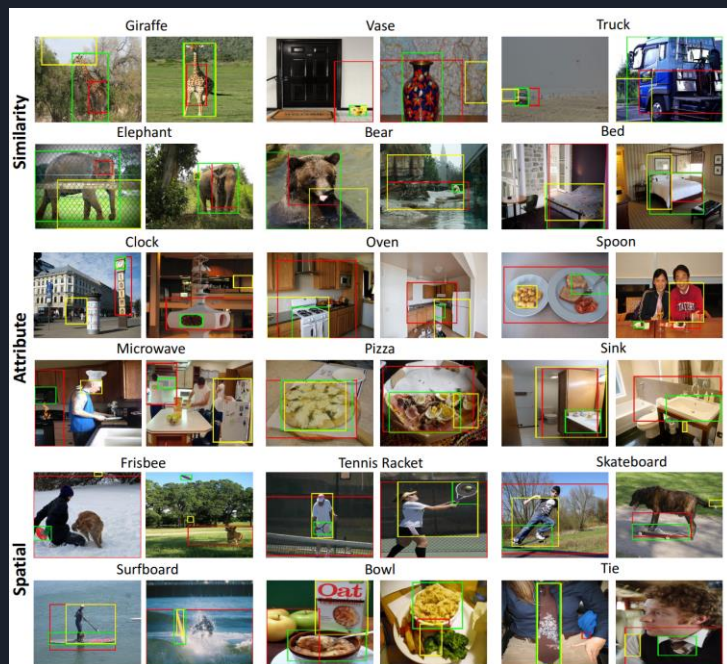




Acquiring common sense

- Class similarity common sense: leverage the semantic similarity of a new target class to previously-learned source classes.
 - Eg. Horse is semantically similar to zebra
 - Compute cosine similarity between feature representations
- Attribute common sense: Mid level semantic visual concepts(furry, red, round)
 - use pre-trained set of attribute classifiers from the ImageNet Attribute knowledge base and object-attribute relationships from the Visual Genome knowledge base
- Spatial common sense: a 'bowl' on a 'table', a 'backpack' is behind a 'person'
 - utilize the information about relative locations and sizes of source object classes from visual genome database
- Scene common sense: a 'surfboard' is more likely to be found on a beach
 - Use SceneUNderstanding (SUN) and Places knowledge bases

Results



Failure cases

The approach fails when the object-of-interest is hardly-visible ('handbag') or when source objects with similar attribute (metallic) are cluttered together ('spoon'). For 'wine glass', the bottle is falsely detected because during training it was provided the common-sense that wine-glass is semantically similar to a bottle.

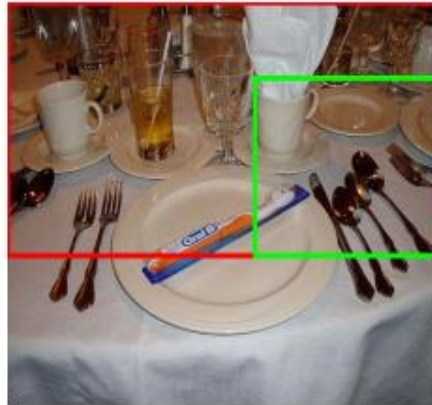
Handbag



Baseball Bat



Spoon



Wine Glass





Paper 2

Deep Image Prior

Author(s): Dmitry Ulyanov et al; Published at CVPR 2018

Additional References: https://dmitryulyanov.github.io/deep_image_prior



Introduction

- The structure of a generator network is sufficient to capture a great deal of low-level image statistics prior to any learning
- A randomly-initialized neural network can be used as a handcrafted prior
- Useful for inverse problems like denoising and inpainting
- For a network to be robust, the structure of the network needs to “resonate” with the structure of the data
- Main contribution of paper: The authors show that contrary to the belief that learning is necessary for building good image priors, a great deal of image statistics are captured by the structure of a convolutional image generator independent of learning.
- Useful to solve image restoration problems, where the image prior is required to integrate information lost in the degradation processes.
- No aspect of the network is learned from the data, randomly initialised weights
- Unsupervised learning approach
- Single Image Approach; No training required, only testing.



Method

- Goal: recover original image x from corrupted image x'
- Formulation: $\min \{E(x;x') + R(x)\}$ over x
- $E(x;x')$ is the data term; L2 norm, KL divergence can be used
- $R(x)$ is the image prior; hard to capture, many methods use deep NN and train using a huge training set
- Instead of searching for the answer in the image space, search for it in the space of neural network's parameters.
- No pre trained network or image database was used; NN was randomly initialised
- Train on a large number of epochs, good hyperparameter tuning

Results

JPEG Artifacts removal



Corrupted



Deep image prior

Inpainting



Corrupted



Deep image prior

Inpainting



Corrupted



Deep image prior

Denoising



Corrupted



Deep image prior



Thank You!