SIMCO: SIMilarity-based object COunting

Marco Godi*, Christian Joppi*, Andrea Giachetti, Marco Cristani
University of Verona

ICPR 2020

* equal contribution

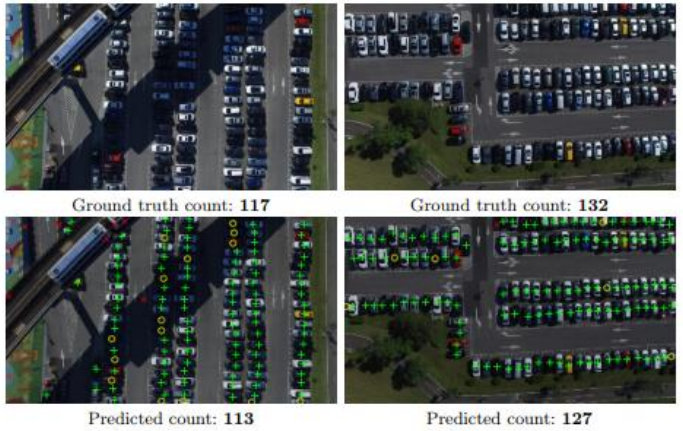
Introduction

- Almost all existing counting methods are designed for a specific object class (crowd counting[1,2], etc..)
- Weaknesses:
  - Need to train each class we want to deal with
  - The classes must be known priorly
  - Huge datasets are necessary for a proper training phase

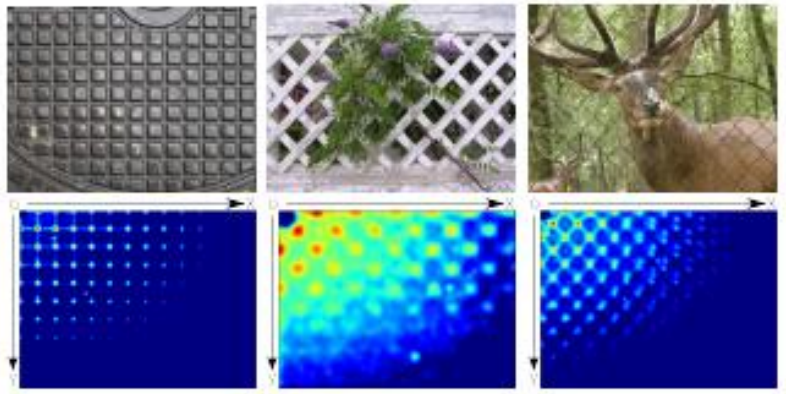
- Class-Agnostic Counting methods ignore these issues and aim to deal with many unknown classes.

State-of-the-Art

- Class-Agnostic Counting[3]
  - not completely agnostic, a form of few-shot learning approach is used for each class

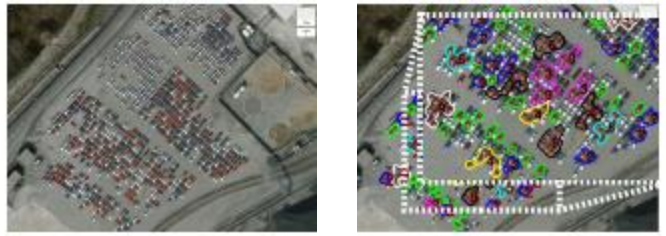


- Repeated Pattern Detection using CNN activations[4]
  - It can detect **any** kind of repeated elements without learning, but only in a regular layout

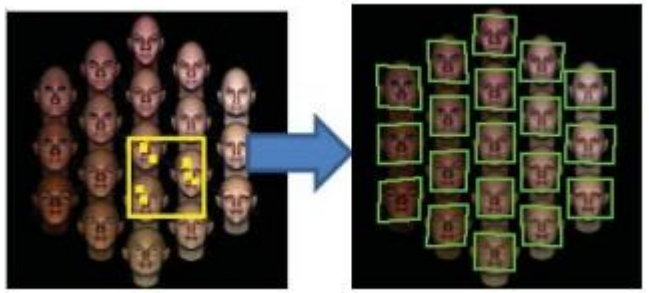


State-of-the-Art

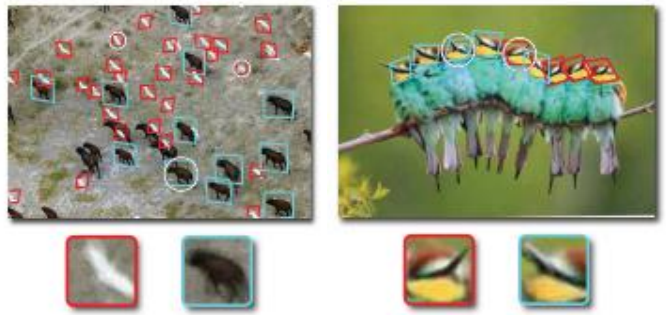
- Interactive Object Counting[5]
  - Completely class-agnostic and without learning
  - Needs user-initialization.



- Detecting, grouping, and structure inference for invariant repetitive patterns in images[6]
  - It is completely agnostic and without learning
  - User or automatic initialization.

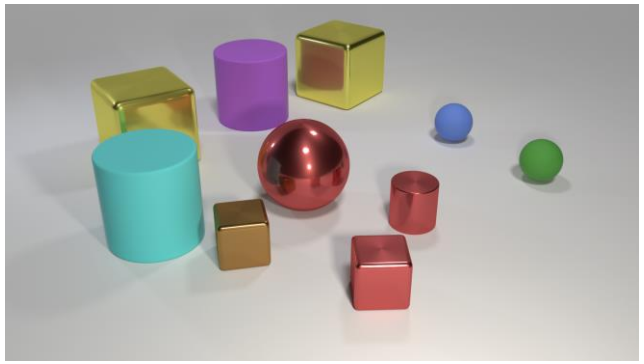


- Count on Me: Learning to Count on a Single Image[7]
  - It is an improved version of the previous work[6].

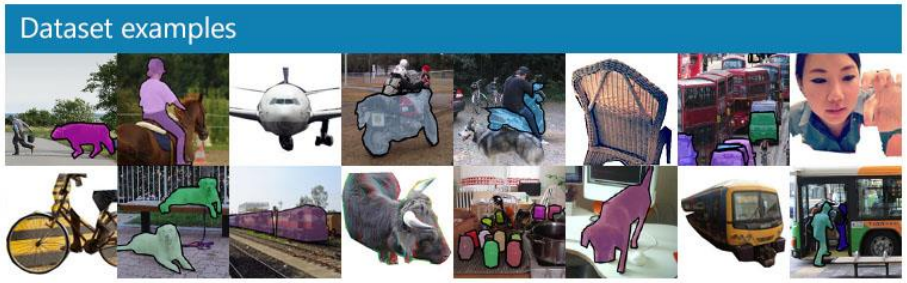


Dataset

- CLEVR[8]: a dataset formed by elementary geometric volumes



- COCO Count[9]: a large-scale object detection, segmentation, and captioning dataset.

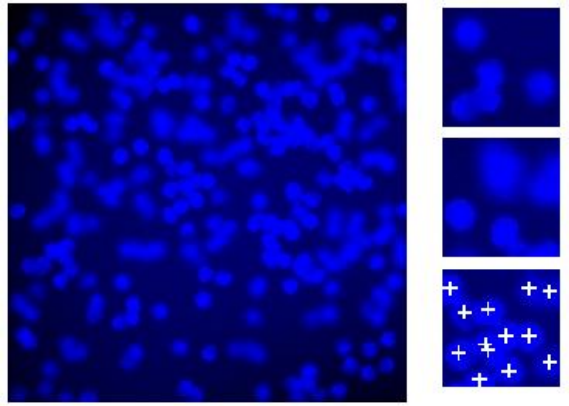


These datasets are not suitable for class-agnostic object counting task since the objects in the image are **rarely repeated** and they belong to **a fixed set of classes**.

Dataset

- Reptile [7]: composed of 50 heterogeneous images taken at different scales, illumination conditions

- Cells[10]: composed by images of a single class of cells in challenging spatial configurations (variable density, occlusions)



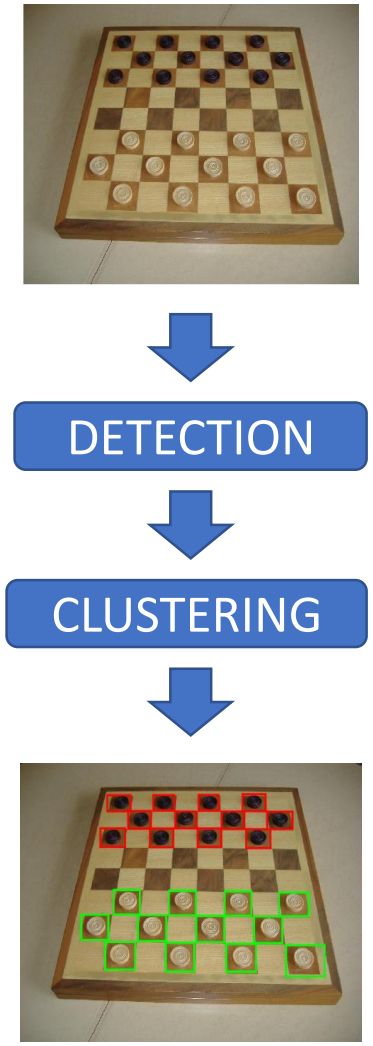
The SIMCO approach

- SIMCO is the first multi-class counting by detection approach, trained just once.

- It is an algorithm composed by two main steps:

  - SIMCO detection provides the bounding box, a single class of generic foreground object and a trained embedding for each detection.

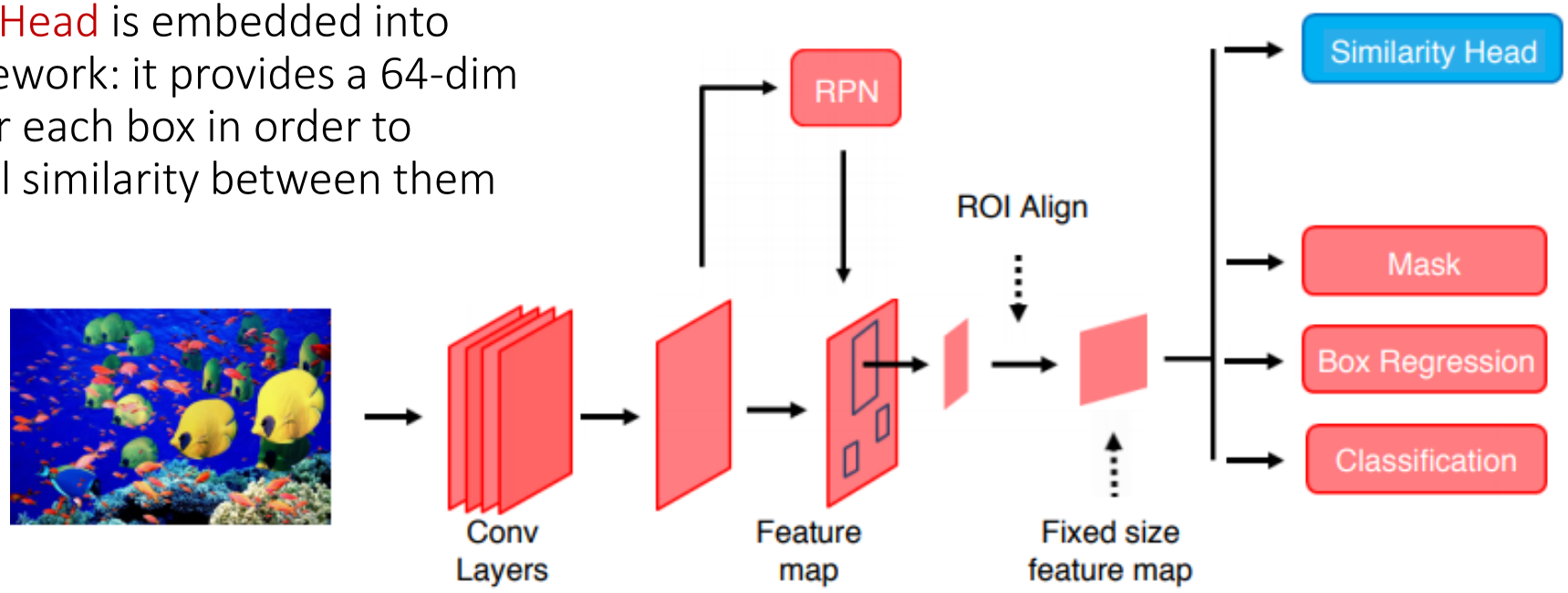
  - SIMCO clustering groups the detection into clusters of different visual «things» using the embedding previously computed



SIMCO Detection

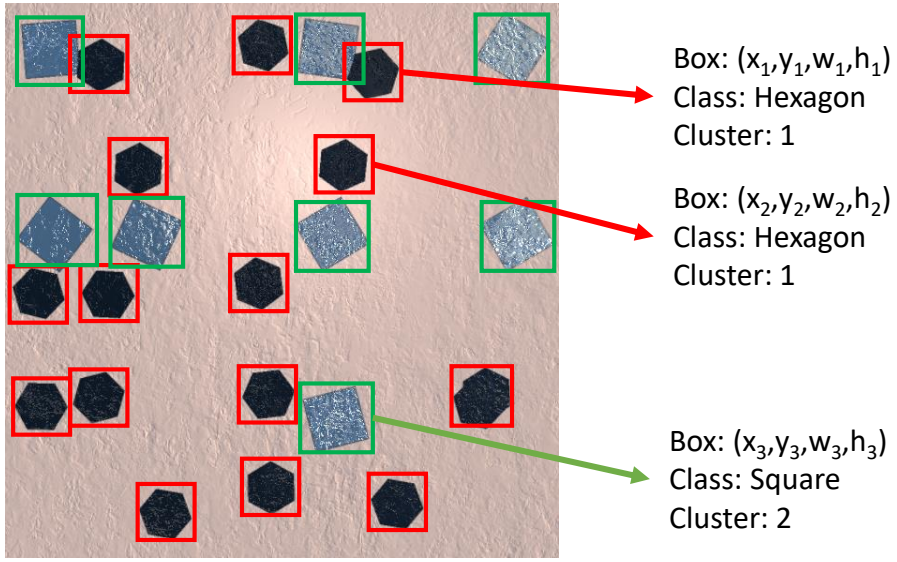
SIMCO builds upon the Mask-RCNN[11] architecture

- A novel Similarity Head is embedded into Mask-RCNN framework: it provides a 64-dim features vector for each box in order to discriminate visual similarity between them

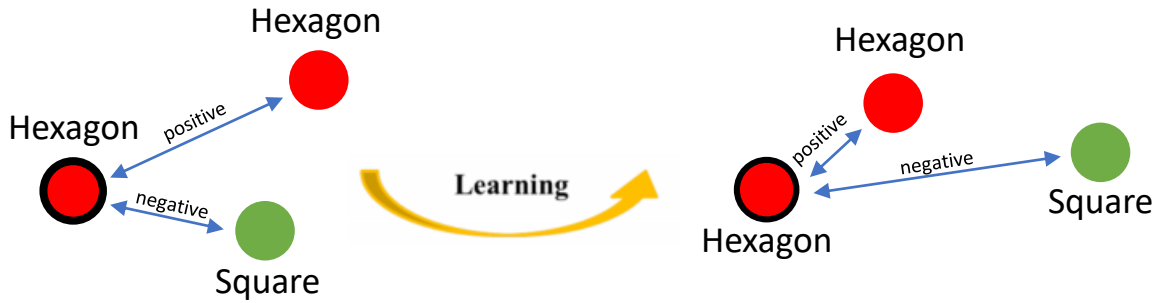


Training on InShape

- INTUITION: each object is directly derived by a primitive 2D shape



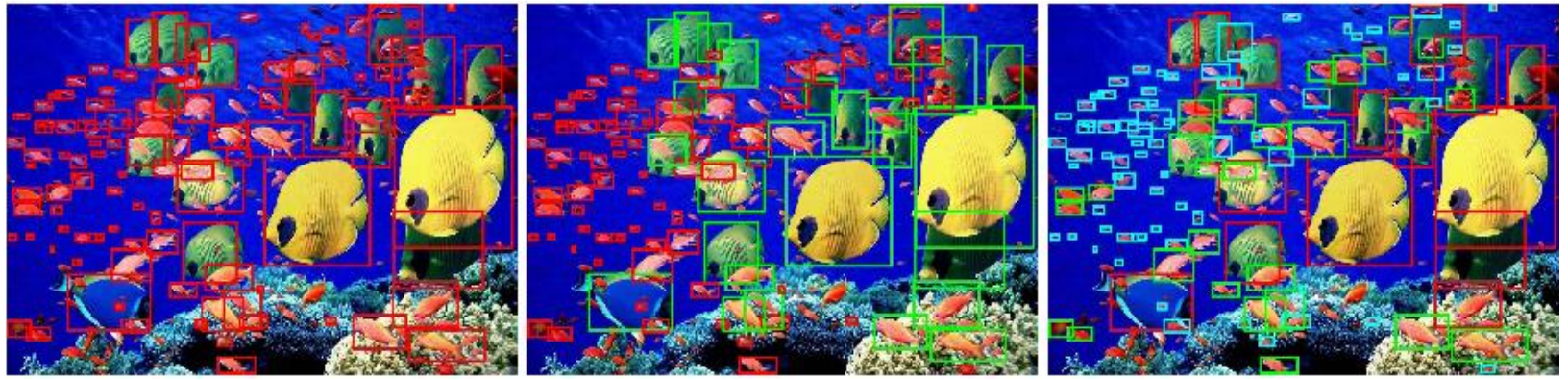
- The model is trained on InShape, a novel dataset of 2D primitive shapes



- Similarity Head has been trained with Triplet Loss function[12], making SIMCO able to discriminate visual similarities between objects (boxes)

SIMCO Clustering

- Starting from the embeddings provided by SIMCO Detection, SIMCO Clustering groups the boxes into «visual things».
  - As clustering procedure we choose the affinity propagation algorithm[13]
  - The single parameter «preference» of the clustering algorithm regulates the tendency to select less or more exemplars



Affinity propagation preference

Experiments

- SIMCO trained on InShape has been tested on two different datasets:
  - Reptile dataset: more than one cluster (semi-automatic protocol)
  - On Cells dataset: one single cluster (completely automatic protocol)
- Counting performance are evaluated using standard index Mean Absolute Error (MAE) and Normalized Mean Absolute Error (NMAE)

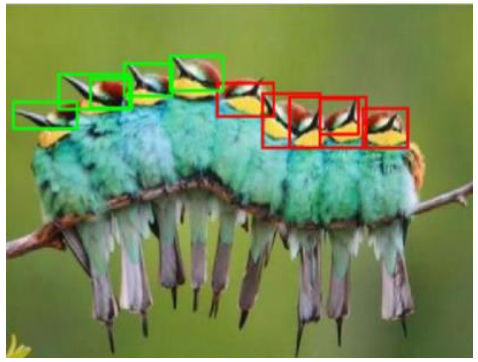
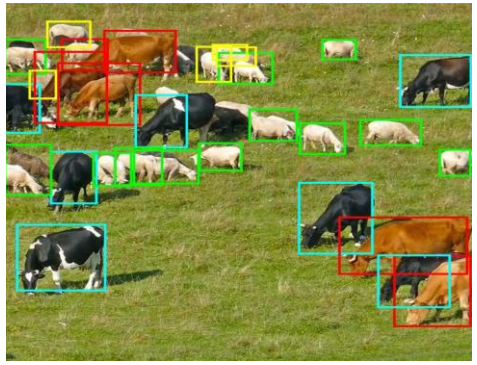
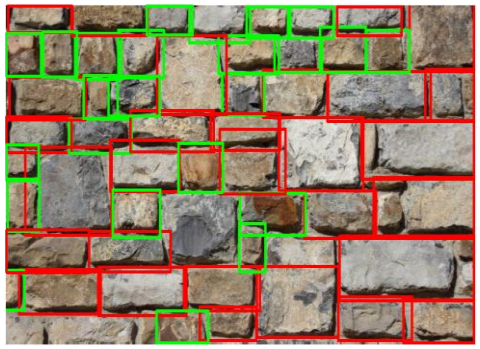
$$MAE = \frac{\sum_{i=1}^{n} |y_i - x_i|}{n} \qquad NMAE = \frac{\sum_{i=1}^{n} |y_i - x_i|}{\sum_{i=1}^{n} x_i}$$

Results

Agnostic Counting Results on Reptile [7]			
Method	**Counting**		**Running Time (s)**
	MAE	**NMAE**	
Cai and Baciu [3]	59	1,034	2814
Arteta et al. [4]	50	1,629	685
Setti et al. TM	18	0,186	-
Setti et al. TM + CE	18	0,164	-
Setti et al. complete [2]	14	0,109	867
COCO/Mask-RCNN/FC	46	0,521	**0,18**
InShape/Mask-RCNN/FC	19	0,272	**0,18**
SIMCO	**8,66**	**0,086**	**0,18**

Agnostic Counting Results on Cells [10]			
Method	**Counting**		**Running Time (s)**
	MAE	**NMAE**	
Cai and Baciu [3]	149	0,809	753
SharpMask [7]	42	0,21	8,76
COCO/Mask-RCNN	175,65	0,99	**0,12**
SIMCO	**12**	**0,07**	**0,11**

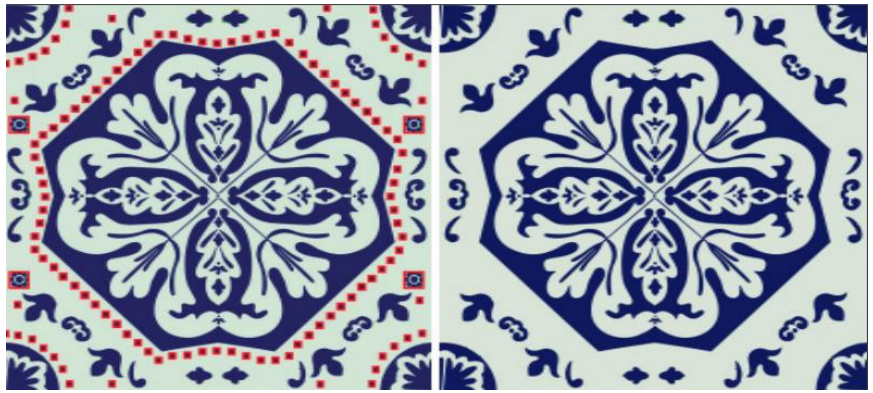



Conclusions

- We presented SIMCO, the first completely class-agnostic counting approach

- Possibile applications:
  - Photoediting
  - Annotation Toolkit



References

1. Xiong, Haipeng, et al. "From open set to closed set: Counting objects by spatial divide-and-conquer." *Proceedings of the IEEE International Conference on Computer Vision*. 2019.
2. Liu, Weizhe, Mathieu Salzmann, and Pascal Fua. "Context-aware crowd counting." *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 2019.
3. Lu, Erika, Weidi Xie, and Andrew Zisserman. "Class-agnostic counting." *Asian conference on computer vision*. Springer, Cham, 2018.
4. Lettry, Louis, et al. "Repeated pattern detection using CNN activations." *2017 IEEE Winter Conference on Applications of Computer Vision (WACV)*. IEEE, 2017.
5. Arteta, Carlos, et al. "Interactive object counting." *European conference on computer vision*. Springer, Cham, 2014.
6. Cai, Yunliang, and George Baciu. "Detecting, grouping, and structure inference for invariant repetitive patterns in images." *IEEE Transactions on Image Processing* 22.6 (2013): 2343-2355.
7. Setti, Francesco, et al. "Count on me: learning to count on a single image." *IEEE Transactions on Circuits and Systems for Video Technology* 28.8 (2017): 1798-1806.
8. Johnson, Justin, et al. "Clevr: A diagnostic dataset for compositional language and elementary visual reasoning." *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 2017.
9. Lin, Tsung-Yi, et al. "Microsoft coco: Common objects in context." *European conference on computer vision*. Springer, Cham, 2014.
10. V. Lempitsky and A. Zisserman, “Learning to count objects in images,” in Advances in neural information processing systems, 2010, pp. 1324– 1332.
11. K. He, G. Gkioxari, P. Dollar, and R. Girshick, “Mask r-cnn,” in ´ International Conference on Computer Vision. IEEE, 2017, pp. 2980– 2988.
12. Schroff, Florian, Dmitry Kalenichenko, and James Philbin. "Facenet: A unified embedding for face recognition and clustering." *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2015.
13. B. J. Frey and D. Dueck, “Clustering by passing messages between data points,” science, vol. 315, no. 5814, pp. 972–976, 2007.