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

[^3]: ICPR 2020
[^4]: ICPR 2020
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\textsuperscript{[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\textsuperscript{[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)

\[
MAE = \frac{\sum_{i=1}^{n} |y_i - x_i|}{n}
\]
\[
NMAE = \frac{\sum_{i=1}^{n} |y_i - x_i|}{\sum_{i=1}^{n} x_i}
\]
## Results

### Agnostic Counting Results on Reptile \(^7\)

<table>
<thead>
<tr>
<th>Method</th>
<th>Counting</th>
<th>Running Time (s)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>MAE</td>
<td>NMAE</td>
</tr>
<tr>
<td>Cai and Baciu [3]</td>
<td>59</td>
<td>1,034</td>
</tr>
<tr>
<td>Arteta et al. [4]</td>
<td>50</td>
<td>1,629</td>
</tr>
<tr>
<td>Setti et al. TM</td>
<td>18</td>
<td>0,186</td>
</tr>
<tr>
<td>Setti et al. TM + CE</td>
<td>18</td>
<td>0,164</td>
</tr>
<tr>
<td>Setti et al. complete [2]</td>
<td>14</td>
<td>0,109</td>
</tr>
<tr>
<td>COCO/Mask-RCNN/FC</td>
<td>46</td>
<td>0,521</td>
</tr>
<tr>
<td>InShape/Mask-RCNN/FC</td>
<td>19</td>
<td>0,272</td>
</tr>
<tr>
<td>SIMCO</td>
<td>8,66</td>
<td>0,086</td>
</tr>
</tbody>
</table>

### Agnostic Counting Results on Cells \(^10\)

<table>
<thead>
<tr>
<th>Method</th>
<th>Counting</th>
<th>Running Time (s)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>MAE</td>
<td>NMAE</td>
</tr>
<tr>
<td>Cai and Baciu [3]</td>
<td>149</td>
<td>0,809</td>
</tr>
<tr>
<td>SharpMask [7]</td>
<td>42</td>
<td>0,21</td>
</tr>
<tr>
<td>COCO/Mask-RCNN</td>
<td>175,65</td>
<td>0,99</td>
</tr>
<tr>
<td>SIMCO</td>
<td>12</td>
<td>0,07</td>
</tr>
</tbody>
</table>

---

ICPR 2020
Conclusions

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

• Possible applications:
  
  • Photoediting
  
  • Annotation Toolkit
References


