GMNet: Graph Matching Network for Large Scale Part Semantic Segmentation in the Wild

Umberto Michieli, Edoardo Borsato, Luca Rossi, Pietro Zanuttigh

umberto.michieli@dei.unipd.it
Semantic Segmentation - Definition

Assign to each pixel a label representing the class to which the pixel belongs.

- Dense task
- Deep learning revolutionized the field (autoencoder models) [1]

Multi-Class Part Parsing

→ Learn multiple parts of multiple objects

Input image

Object-level parsing

Single-class part parsing (e.g. person)

Multi-class part parsing

58 parts

108 parts
Coarse-to-Fine Learning

Transfer knowledge from a coarse problem to a finer one

**Spatial level** coarse-to-fine: object-level classes split into their parts

→ learn multiple parts of multiple objects
Coarse-to-Fine at Spatial Level

First idea (baseline): just train a network on all the different parts

Low results, 2 main reasons:

- **Object-level ambiguity**: corresponding parts in different semantic classes often share similar appearance

![Sheep legs](image1)

? 

![Cow legs](image2)
Coarse-to-Fine at Spatial Level

First idea (baseline): just train a network on all the different parts

Low results, 2 main reasons:

- **Object-level ambiguity:** corresponding parts in different semantic classes often share similar appearance
Coarse-to-Fine at Spatial Level

First idea (baseline): just train a network on all the different parts

Low results, 2 main reasons:

- **Object-level ambiguity**: corresponding parts in different semantic classes often share similar appearance

- **Part-level ambiguity**: limited local context is captured
Coarse-to-Fine at Spatial Level

First idea (baseline): just train a network on all the different parts

Low results, 2 main reasons:

- **Object-level ambiguity**: corresponding parts in different semantic classes often share similar appearance

- **Part-level ambiguity**: limited local context is captured
Coarse-to-Fine at Spatial Level

First idea (baseline): just train a network on all the different parts

Low results, 2 main reasons:

- **Object-level ambiguity:** corresponding parts in different semantic classes often share similar appearance
  - object-level guidance via semantic embedding network $S$
  - auxiliary reconstruction module from parts to objects

- **Part-level ambiguity:** limited local context is captured
  - graph-matching module to preserve relative spatial relationships between ground truth and predicted parts.
GMNet Architecture

Channel-wise concatenation

Trainable

Pre-trained on object parsing
Normalized matrices → proximity ratios

Graph-Matching loss:
\[ \mathcal{L}_{GM} = \| \mathbf{M}^{GT} - \mathbf{M}^{pred} \|_F \]
Dataset – VOC2012 Pascal Parts

RGB

Object-level GT

Pascal-Part-58

Pascal-Part-108

PASCAL-VOC 2012:

- 10103 images: 4998 train and 5105 validation
- 21 object-level classes
- Pascal-Part-58 [1] and Pascal-Part-108 [2,3]

Experiments – Pascal 58

<table>
<thead>
<tr>
<th>Method</th>
<th>mIoU</th>
<th>Avg.</th>
</tr>
</thead>
<tbody>
<tr>
<td>SegNet</td>
<td>24.4</td>
<td>26.5</td>
</tr>
<tr>
<td>FCN</td>
<td>42.3</td>
<td>44.9</td>
</tr>
<tr>
<td>DeepLab v1</td>
<td>49.9</td>
<td>51.9</td>
</tr>
<tr>
<td>DRN D 38</td>
<td>50.0</td>
<td>50.9</td>
</tr>
<tr>
<td>DRN D 105</td>
<td>53.0</td>
<td>53.0</td>
</tr>
<tr>
<td>BSANet*</td>
<td>58.2</td>
<td>58.9</td>
</tr>
<tr>
<td>Baseline (DeepLab v3)</td>
<td>54.4</td>
<td>55.7</td>
</tr>
<tr>
<td>GMNet (ours)</td>
<td>59.0</td>
<td>61.8</td>
</tr>
</tbody>
</table>

* It is the only other method for multi-class part parsing and uses the same architecture (DeepLab v3+, ResNet-101)

Multi-class Zhao et al., “Part Parsing with Joint Boundary-Semantic Awareness”, ICCV 2019
Experiments – Pascal 108

<table>
<thead>
<tr>
<th>Method</th>
<th>mIoU</th>
<th>Avg.</th>
</tr>
</thead>
<tbody>
<tr>
<td>SegNet</td>
<td>18.6</td>
<td>20.8</td>
</tr>
<tr>
<td>FCN</td>
<td>31.6</td>
<td>33.8</td>
</tr>
<tr>
<td>DeepLab v1</td>
<td>35.7</td>
<td>40.8</td>
</tr>
<tr>
<td>DRN D 38</td>
<td>39.1</td>
<td>41.9</td>
</tr>
<tr>
<td>DRN D 105</td>
<td>39.5</td>
<td>41.0</td>
</tr>
<tr>
<td>BSANet*</td>
<td>42.9</td>
<td>46.3</td>
</tr>
<tr>
<td>Baseline (DeepLab v3)</td>
<td>41.3</td>
<td>43.7</td>
</tr>
<tr>
<td>GMNet (ours)</td>
<td>45.8</td>
<td>50.5</td>
</tr>
</tbody>
</table>

* It is the only other method for multi-class part parsing and uses the same architecture (DeepLab v3+, ResNet-101)

Multi-class Zhao et al., “Part Parsing with Joint Boundary-Semantic Awareness”, ICCV 2019
Conclusion

Semantic segmentation of multiple parts from multiple objects

Contributions:

• Object-level semantic embedding network guides part-level decoding stage
• Graph-matching module for accurate relative localization of semantic parts
• GMNet achieves new state-of-the-art performance on Pascal-Part-58 and 108
Paper website: https://lttm.dei.unipd.it/paper_data/GMNet

Code: https://github.com/LTTM/GMNet


Contact: umberto.michieli@dei.unipd.it