ECCV'20 ONLINE

23-28 AUGUST 2020

16TH EUROPEAN CONFERENCE ON

COMPUTER VISION

WWW.ECCV2020.EU

ECCV'20 ONLINE
23-28 AUGUST 2020

UNIVERSITÀ DEGLI STUDI DI PADOVA

LTTM

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

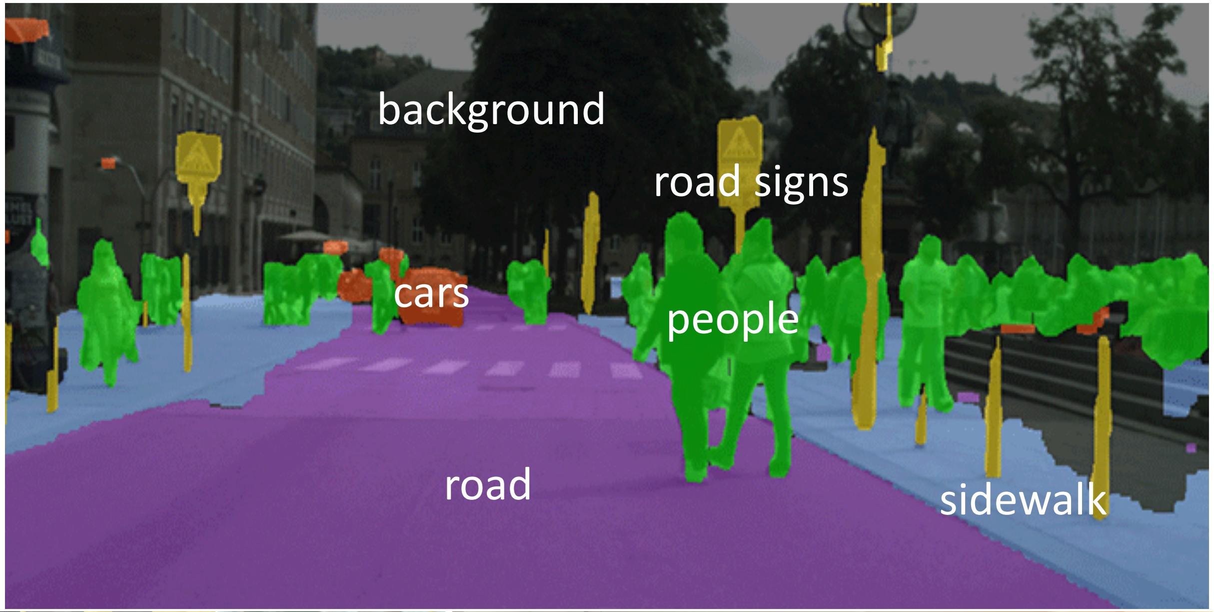
Umberto Michieli, Edoardo Borsato, Luca Rossi, Pietro Zanuttigh

umberto.michieli@dei.unipd.it

Semantic Segmentation - Definition

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

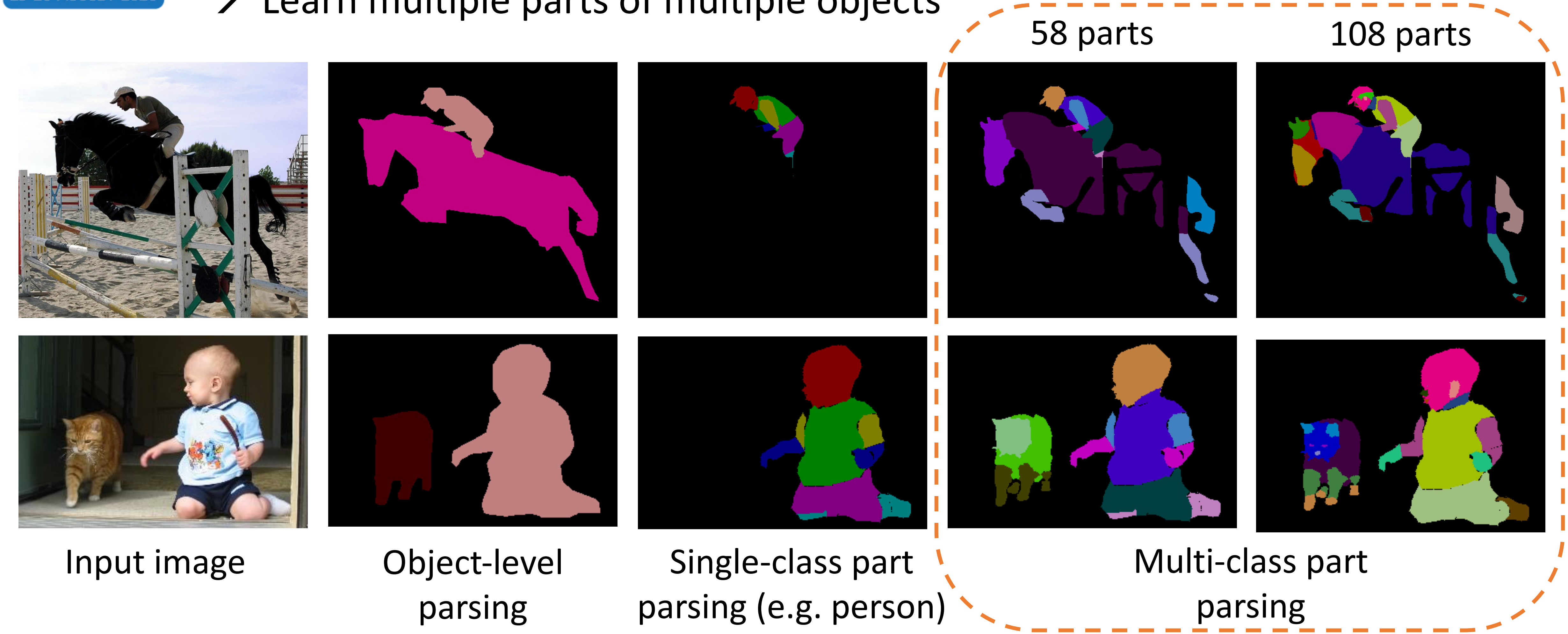
- Dense task
- Deep learning revolutionized the field (autoencoder models) [1]



[1] Long et al., "Fully convolutional networks for semantic segmentation", CVPR 2015.

Multi-Class Part Parsing

→ Learn multiple parts of multiple objects

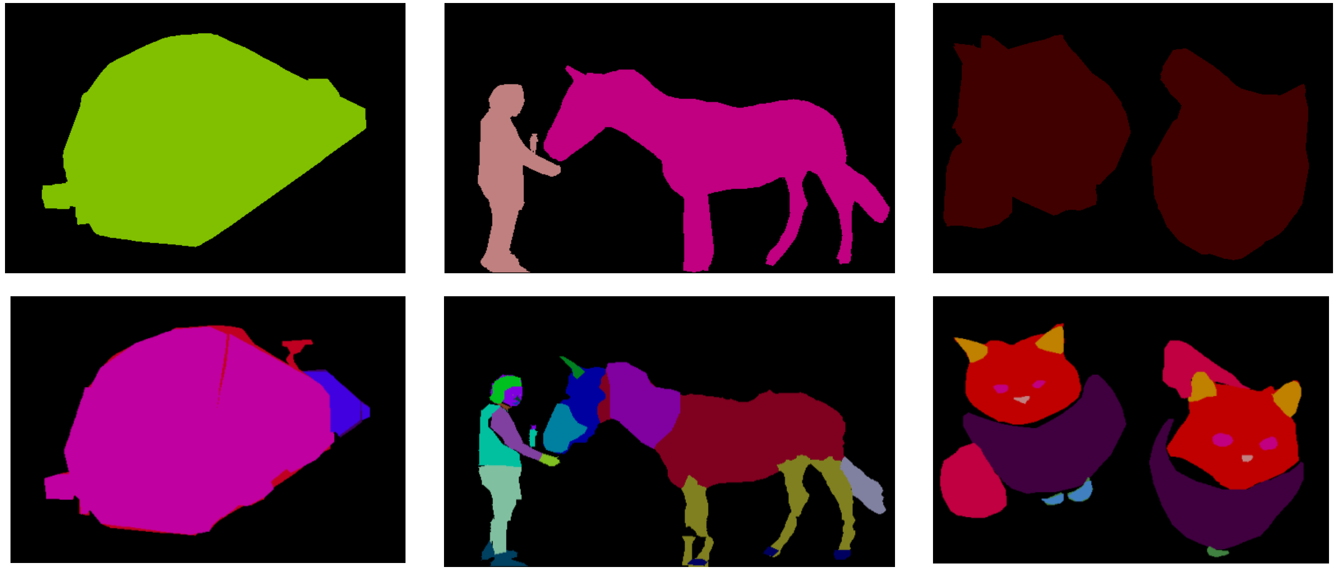


Coarse-to-Fine Learning

Transfer knowledge form 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



Annotations object-level

Annotations part-level

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

?

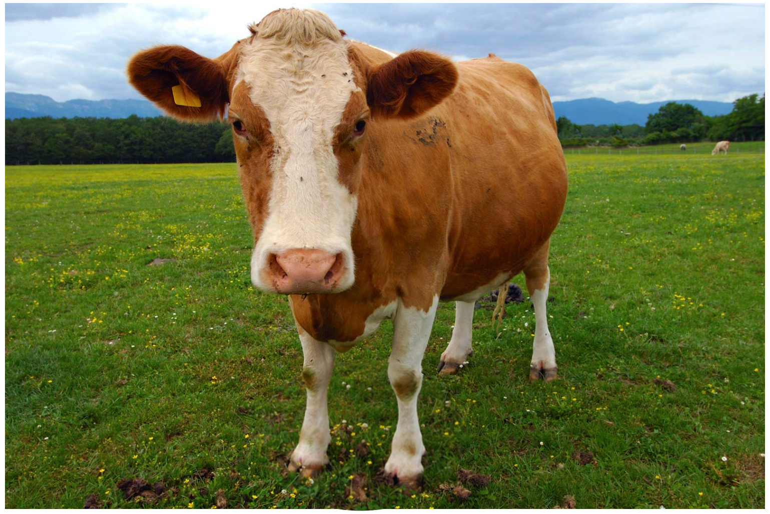
Cow legs

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

?

Cow legs

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

Dog head

?

Dog tail

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

Dog head

?

Dog tail

Coarse-to-Fine at Spatial Level

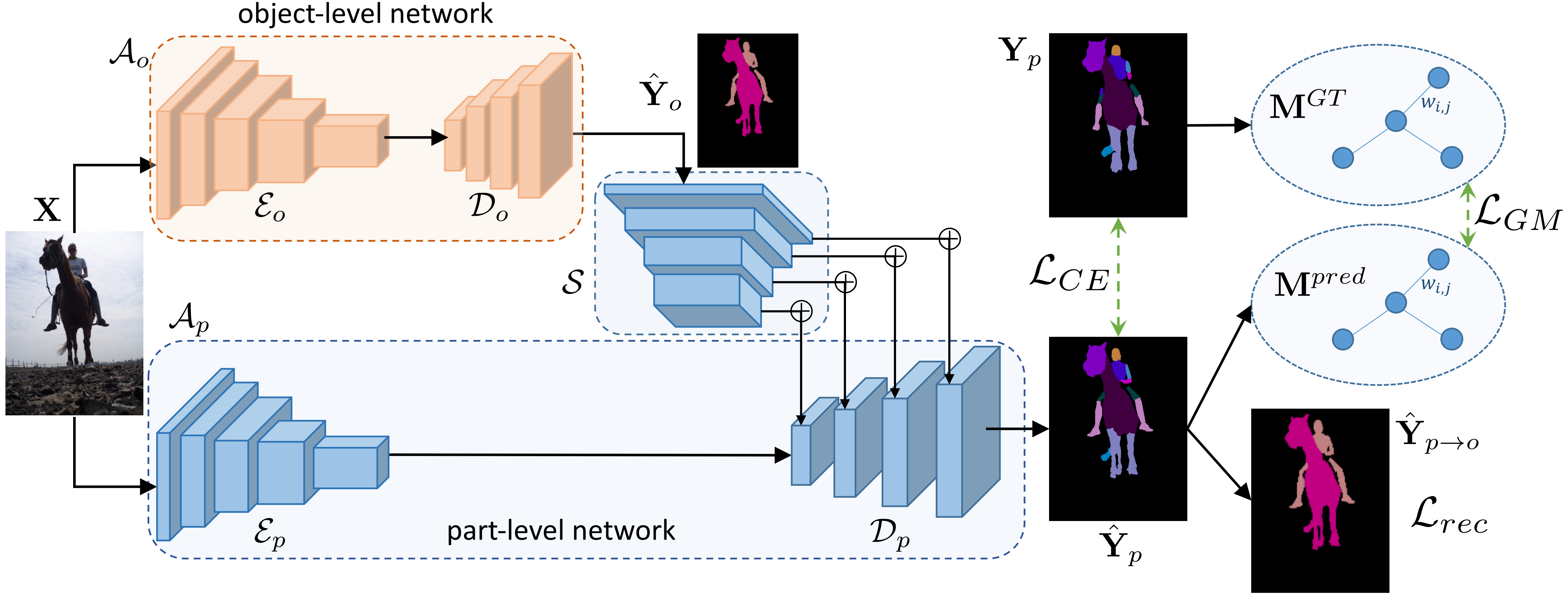
First idea (baseline): just train a network on all the different parts

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 $\mathcal{S}$
  - ➢ auxiliary reconstruction module from parts to objects
- ❑ Part-level ambiguity: limited local context is captured
  - ➢ graph-matching module to preserve relative spatial relationships

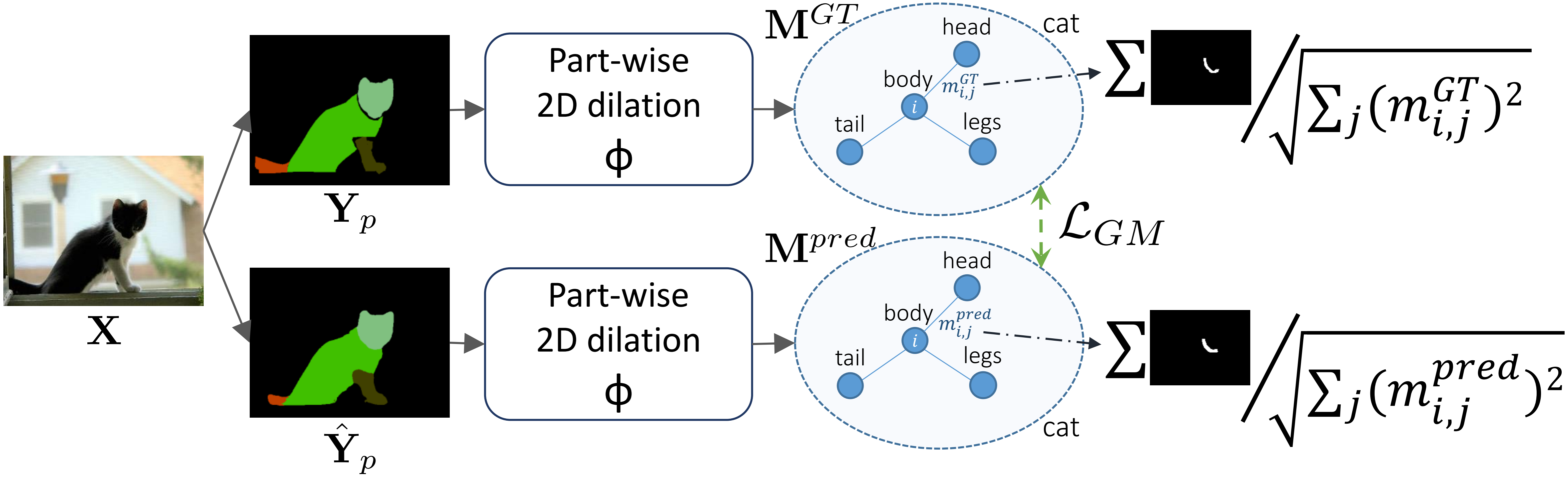
    between ground truth and predicted parts.

GMNet Architecture



$\oplus$ Channel-wise concatenation | Trainable | Pre-trained on object parsing

Graph Matching Module

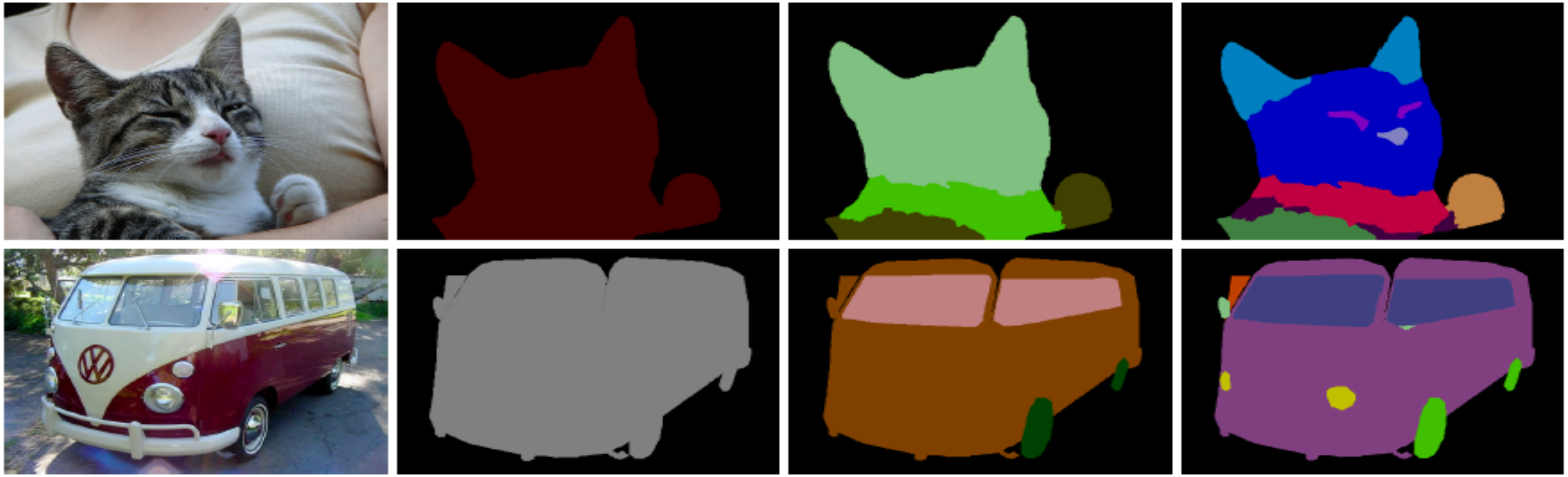


Normalized matrices → *proximity ratios*

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

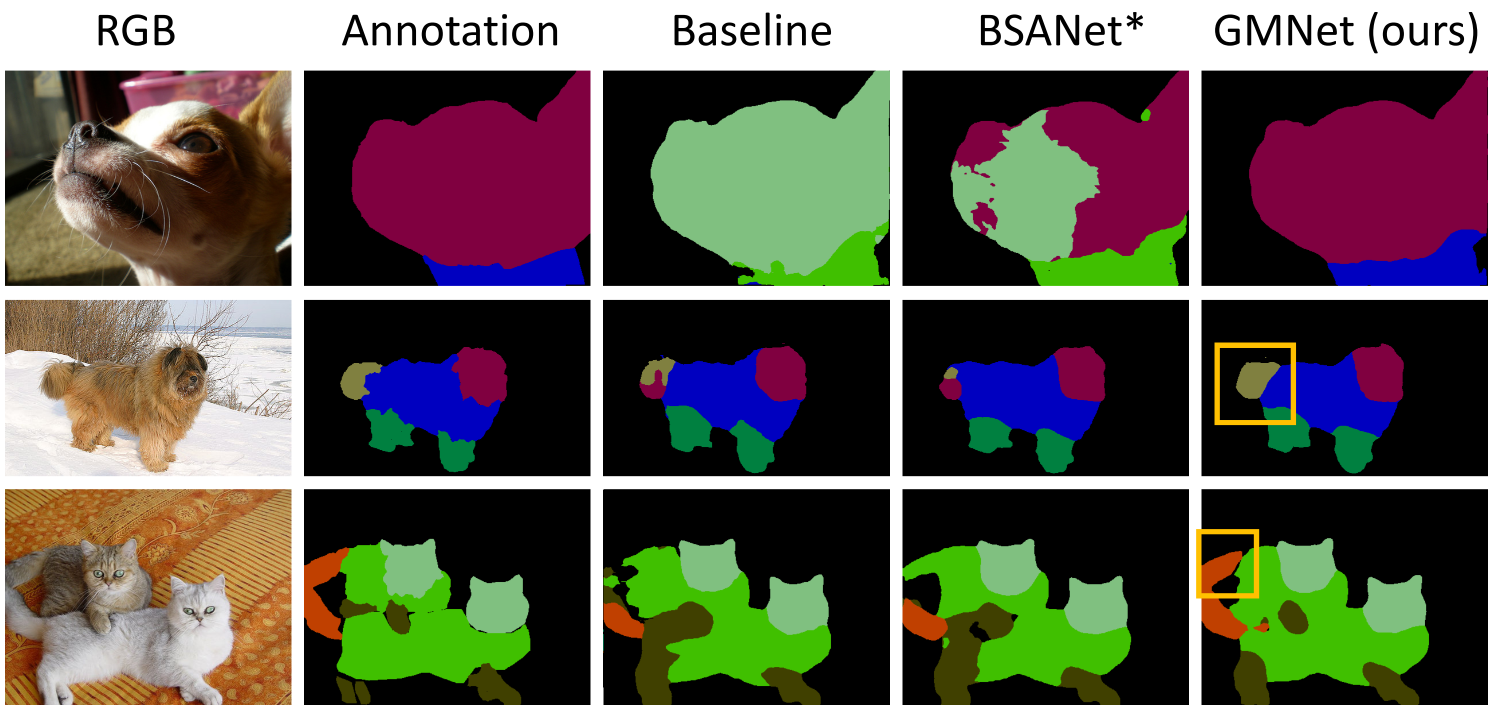


PASCAL-VOC 2012:

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

[1] Zhao et al., "Multi-class Part Parsing with Joint Boundary-Semantic Awareness", iCCV 2019
[2] A. Gonzalez-Garcia et al., "Do Semantic Parts Emerge in Convolutional Neural Networks?", IJCV, 2017
[3] Michieli et al., "GMNet: Graph Matching Network for Large Scale Part Semantic Segmentation in the Wild", ECCV, 2020

Experiments – Pascal 58



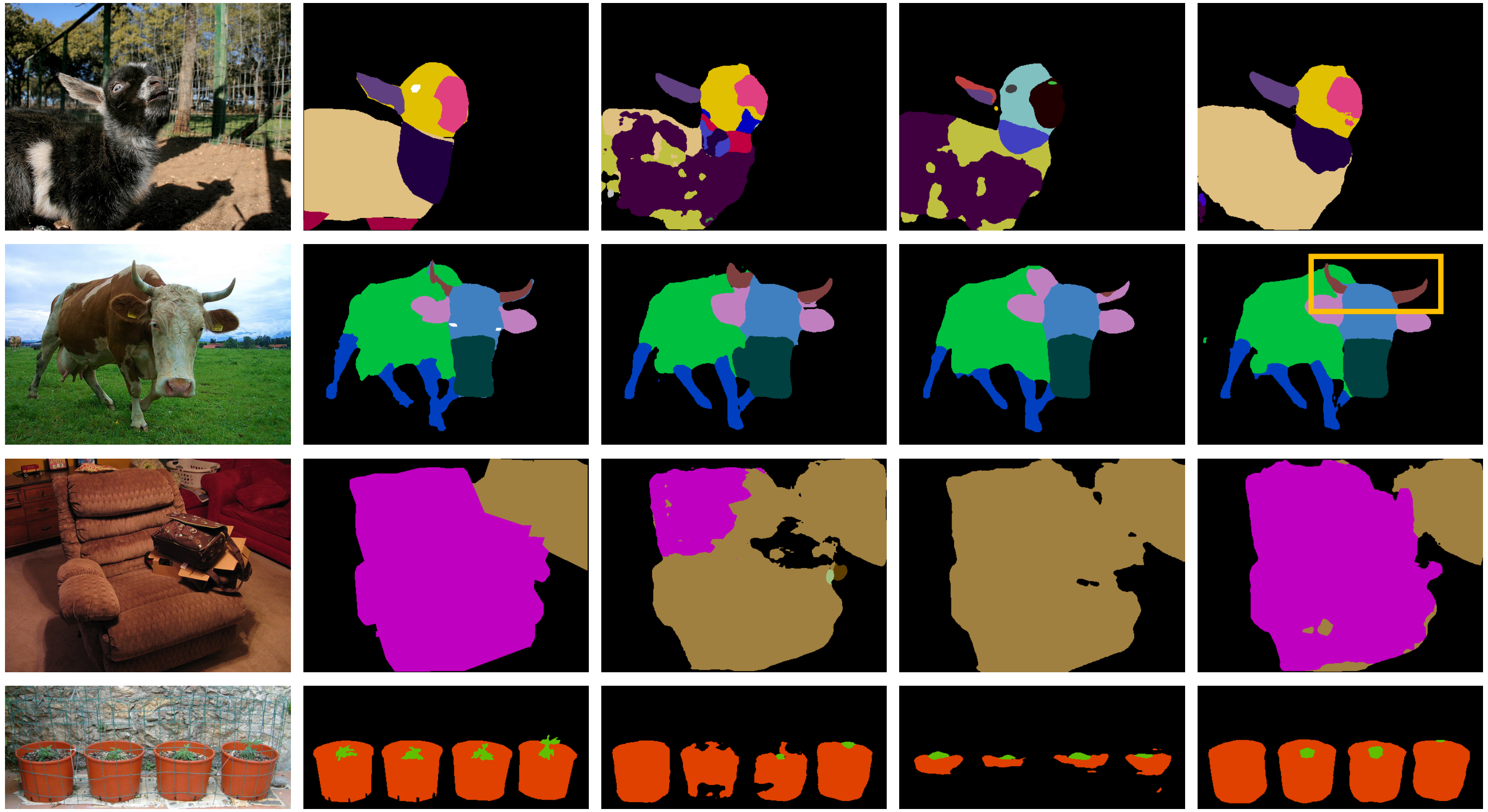
Method	mIoU	Avg.
SegNet	24.4	26.5
FCN	42.3	44.9
DeepLab v1	49.9	51.9
DRN D 38	50.0	50.9
DRN D 105	53.0	53.0
BSANet*	58.2	58.9
Baseline (DeepLab v3)	54.4	55.7
GMNet (ours)	**59.0**	**61.8**

* 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

ECCV'20 ONLINE 23-28 AUGUST 2020

Experiments – Pascal 108



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

* 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

ArXiv: https://arxiv.org/abs/2007.09073

Contact: umberto.michieli@dei.unipd.it

Michieli U., Borsato E., Rossi L. and Zanuttigh P., "GMNet: Graph Matching Network for Large Scale Part Semantic Segmentation in the Wild," ECCV 2020.

ECCV'20 ONLINE
23-28 AUGUST 2020

UNIVERSITÀ DEGLI STUDI DI PADOVA

LTTM