EBERHARD KARLS
UNIVERSITÄT
TÜBINGEN

MATHEMATISCH-
NATURWISSENSCHAFTLICHE FAKULTÄT

FourierNet: Compact Mask Representation for Instance Segmentation Using Differentiable Shape Decoders

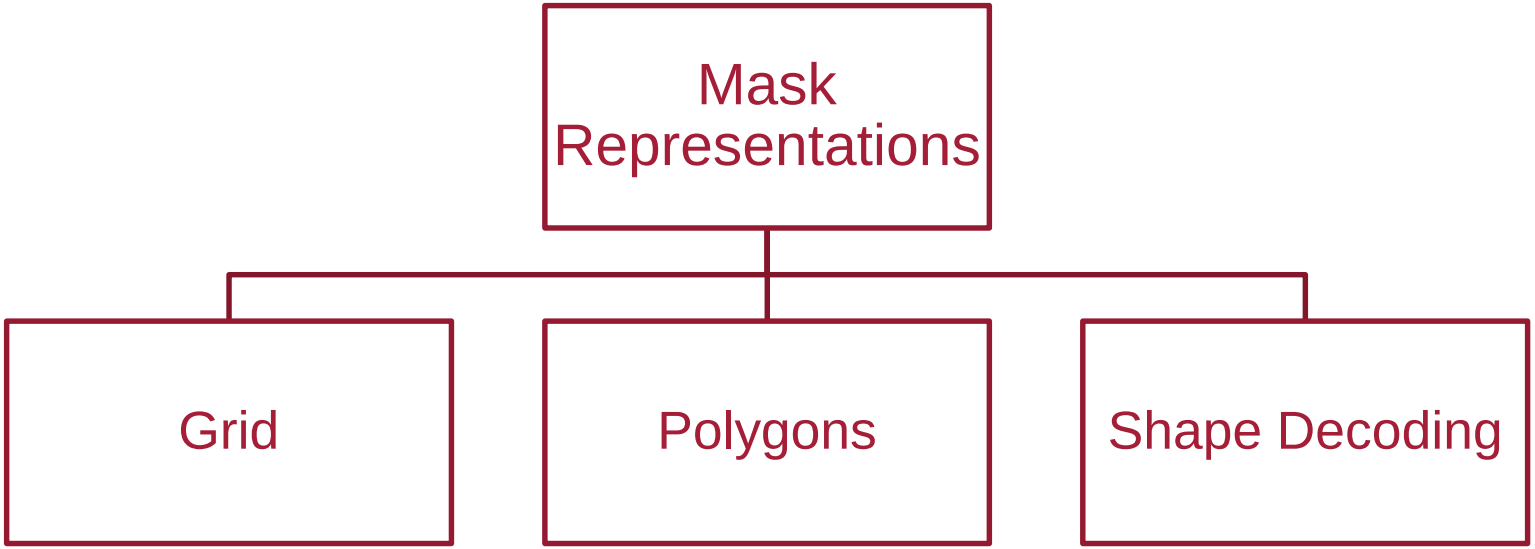
Hamd ul Moqeet Riaz*, Nuri Benbarka* and Andreas Zell

** equal contribution*

Goal

EBERHARD KARLS
UNIVERSITÄT
TÜBINGEN

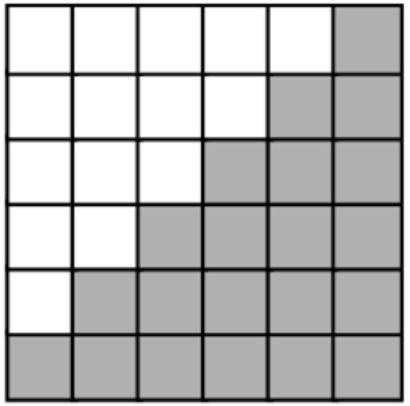
"We want to find a mask representation for segmentation which is compact and informative"

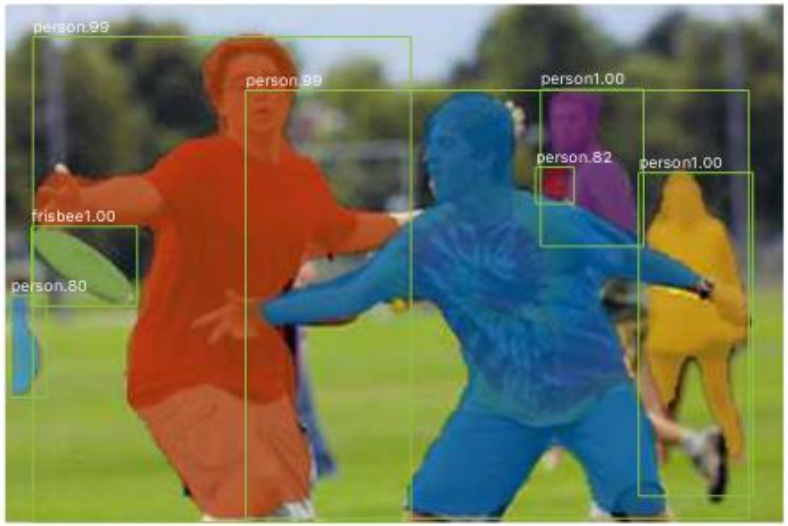


Grid Representation

EBERHARD KARLS
UNIVERSITÄT
TÜBINGEN

- Discretization of the space into grid
- Easy to process with neural networks
- High memory footprint



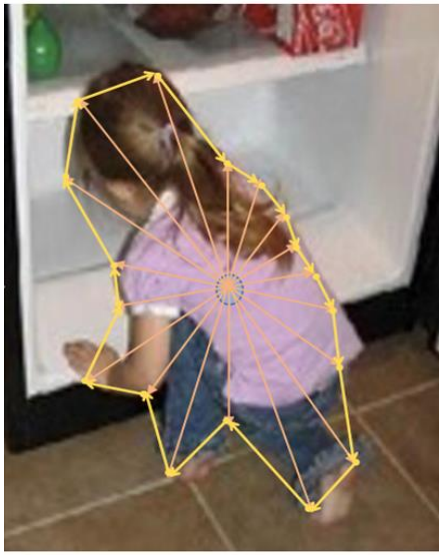


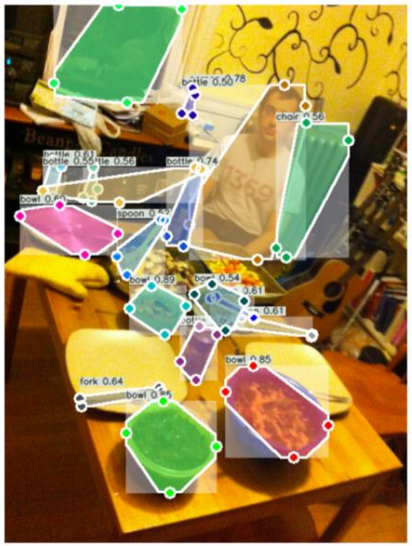
[1] He, Kaiming, et al. "Mask r-cnn." *ICCV* 2017.
[2] Mescheder, Lars, et al. "Occupancy networks: Learning 3d reconstruction in function space." *CVPR* 2019.

Polygons representation

EBERHARD KARLS
UNIVERSITÄT
TÜBINGEN

- Represents the contour as points with a fixed connectivity
- Limited number of points.
- Low memory footprint.



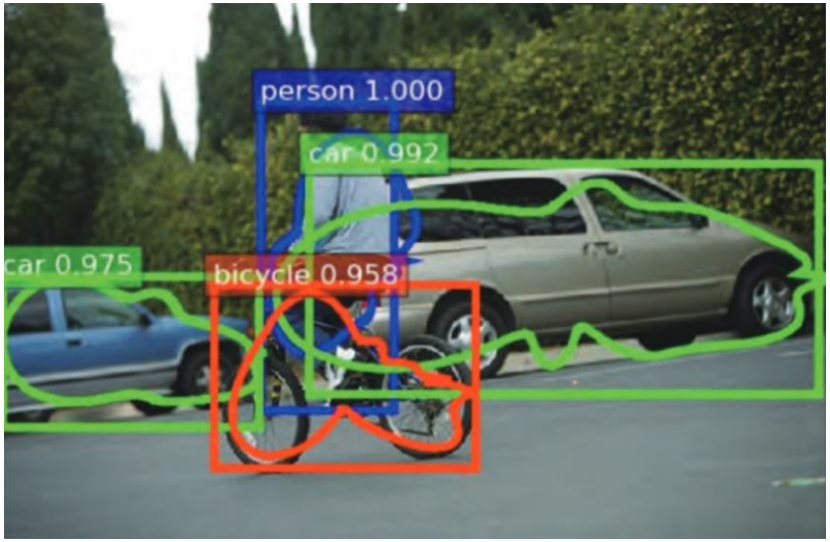


[2] Mescheder, Lars, et al. "Occupancy networks: Learning 3d reconstruction in function space." *CVPR* 2019.
[3] Xie, Enze et al. “PolarMask: Single Shot Instance Segmentation with Polar Representation” *CVPR* 2020
[4] Zhou, Xingyi et al. “Bottom-up Object Detection by Grouping Extreme and Center Points“ CVPR 2019

Shape Decoding

EBERHARD KARLS
UNIVERSITÄT
TÜBINGEN

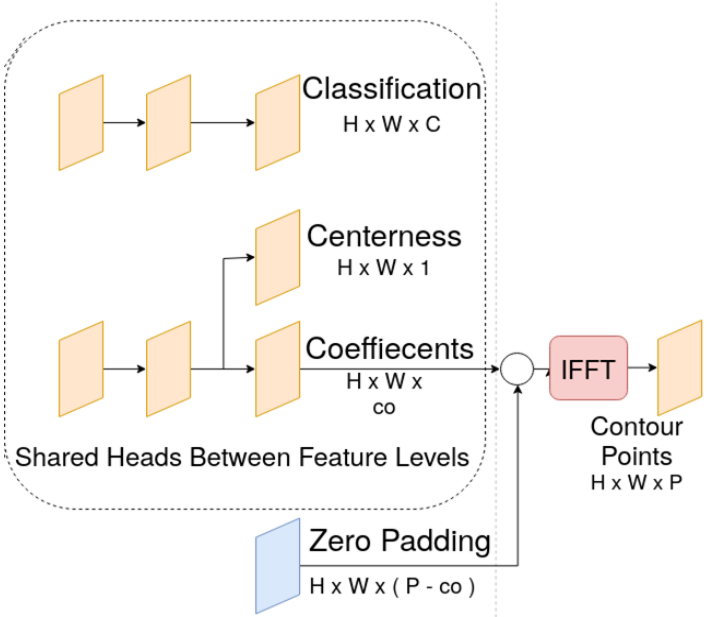
- Decodes a compressed vector into a mask.
- Approximate mask.
- Lower memory footprint

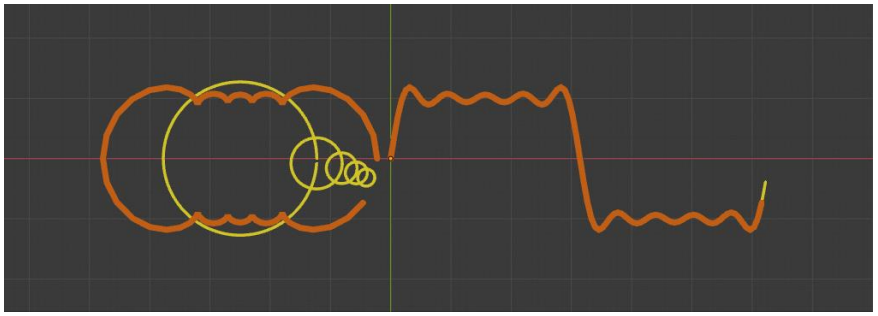


[5] Xu, Wenqiang et al. “Explicit Shape Encoding for Real-Time Instance Segmentation“ ICCV 2019

Our Idea

- Fourier Series as a shape decoder
- Small number of coefficients
- Differentiable
- Monotonically increasing complexity with increasing number of coefficients





[6] https://medium.com/@contra/drawing-anything-with-fourier-series-using-blender-and-python-c0881e1b738c

2 to 36 coefficients

EBERHARD KARLS
UNIVERSITÄT
TÜBINGEN

2 Coefficients

3 Coefficients

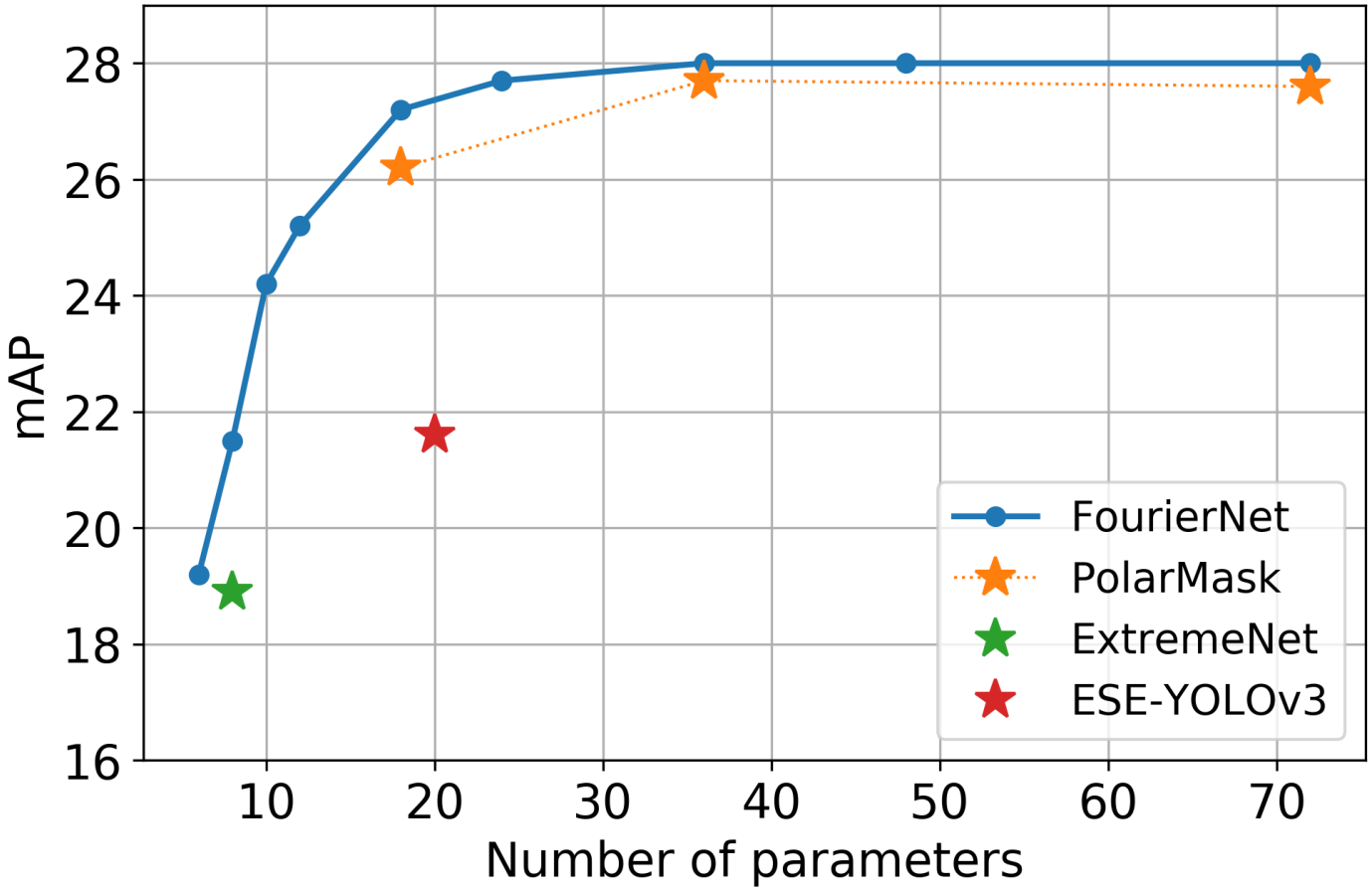
4 Coefficients

6 Coefficients

12 Coefficients

36 Coefficients

Coefficients and contour points



State of the art

EBERHARD KARLS
UNIVERSITÄT
TÜBINGEN

Method	B.Bone	Rep.	Param.	mAP	AP_{50}	AP_{75}	AP_S	AP_M	AP_L	FPS	GPU
two stage											
Mask RCNN [2]	RX-101	binary grid	784	37.1	60.0	39.4	16.9	39.9	53.5	5.6	1080Ti
PANet [3]	RX-101	binary grid	784	**42.0**	**65.1**	**45.7**	22.4	**44.7**	**58.1**	-	-
HTC [5]	RX-101	binary grid	784	41.2	63.9	44.7	**22.8**	43.9	54.6	2.1	TitanXp
one stage											
ESE-Seg-416 [10]	DN-53	shp. encoding	20	21.6	**48.7**	22.4	-	-	-	**38.5**	1080Ti
FourierNet-640	R-50	shp. encoding	20	**24.3**	42.9	**24.4**	6.2	25.9	42.0	26.6	2080Ti
ExtremeNet [12]	HG-104	polygon	**8**	18.9	44.5	13.7	**10.4**	20.4	28.3	3.1	-
FourierNet	RX-101	shp. encoding	**8**	**23.3**	**46.7**	**21.1**	10.3	**25.2**	**34.4**	**6.9**	2080Ti
EmbedMask [11]	R-101	binary grid	†	**37.7**	**59.1**	**40.3**	**17.9**	**40.4**	**53.0**	13.7	V100
YOLACT-700 [8]	R-101	binary grid	†	31.2	50.6	32.8	12.1	33.3	47.1	**23.4**	TitanXp
PolarMask [9]	RX-101	polygon	36	32.9	55.4	33.8	15.5	35.1	46.3	7.1*	2080Ti
FourierNet	RX-101	shp. encoding	36	30.6	50.8	31.8	12.7	33.7	45.2	6.9	2080Ti

Summary

EBERHARD KARLS
UNIVERSITÄT
TÜBINGEN

- FourierNet is a single-stage anchor-free method for instance segmentation.
- We only require small number of coefficients to produce reasonable masks.
- It is a differentiable pipeline and therefore end-to-end trainable.
- FourierNet outperformed all methods which use less than 20 parameters.