Progressive Semantic Segmentation

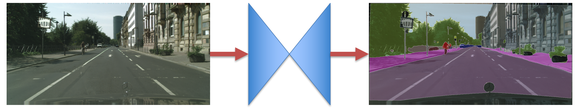
Chuong Huynh[1] Anh Tuan Tran[1,2] Khoa Luu[1,3] Minh Hoai[1,4]

1. VinAI Research 2. Vin University 3. University of Arkansas 4. Stony Brook University

CVPR VIRTUAL JUNE 19-25

Introduction

Semantic segmentation for **High-resolution image**

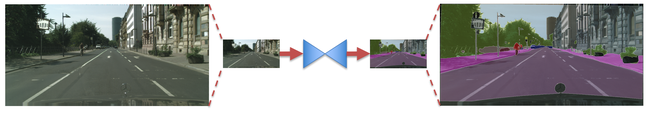


High-resolution image ($\geq 2K \sim 4\ mil\ pixels$) | Segmentation model | Category predictions for each pixel

✗ *Memory overloading when using the original input size*

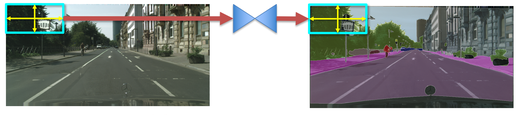
Tradditional approaches

❑ Downsampling



✗ *Missing fine-grained detail*

❑ Patch-based processing:

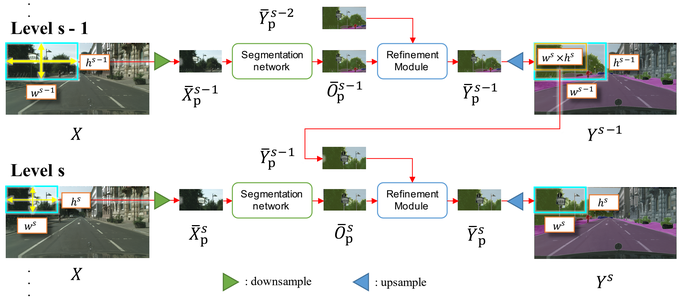


✗ *Missing global context*

MagNet: Combine downsampling and patch-based processing

MagNet

MagNet is an *iterative segmentation* and *refinement* framework



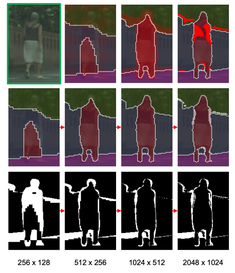
❑ Segmentation maps are constructed **progressively** via **different scales** of input images.

❑ At each scale, inputs are **divided to patches** to remain the **constant memory** usage.

❑ Segmentation maps are corrected **from the coarsest to the finest level**.

❑ Prediction at each scale **contributes fine-grained details** to output from previous scale with two parts:

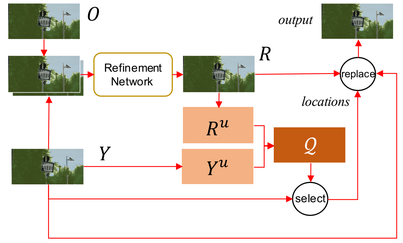
- **Segmentation network:** produces scale-specific prediction.
- **Refinement module:** combine scale-specific prediction with output from previous scale.



Refinement module selectively replaces a part of the previous scale output to:

❑ prevent the output from *"too global"* or *"too local"* to a specific scale.

❑ run faster when it just need to select and refines on a subset of patches (**MagNet-Fast**)



Y : prediction of previous scale

O : scale-specific prediction

R : fused prediction

Query score for ranking new prediction

$$Q = \mathcal{F}(Y^u \odot (1 - R^u))$$

$\odot$: point-wise multiplication

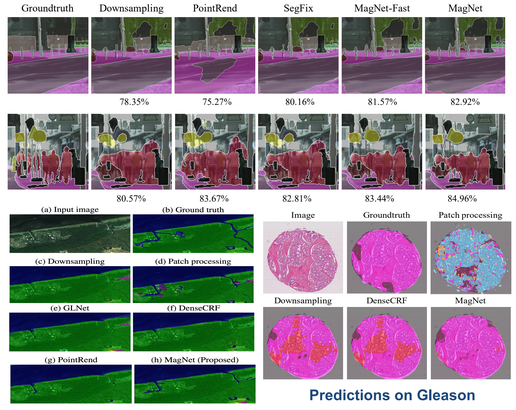
$\mathcal{F}$: median blur

Y^u, R^u : prediction uncertainty

Each stage, refinement module will select k points from Y to replace with R.

Experiments

	Cityscapes [10]: 2048 x 1024			DeepGlobe [11]: 2448 x 2448			Gleason [12]: 5000 x 5000		
	mIoU(%)	Mem. (MB)	Time(s)	mIoU(%)	Mem. (MB)	Time(s)	mIoU(%)	Mem. (MB)	Time(s)
Patching	52.19	1575	1.77	70.98	1247	0.31	46.56	1903	2.42
Downsample	63.23	1575	0.02	67.86	1247	0.01	68.90	1903	0.02
DenseCRF [1]	62.95	1575	26.02	70.36	1247	39.68	69.46	1903	141.79
DGF [3]	63.33	1727	0.32	70.38	1435	0.25	68.91	2223	0.29
GLNet [5]	-	-	-	71.60	1865	0.37	-	-	-
PointRend [7]	64.39	2033	0.14	71.78	1593	0.16	68.97	2655	0.21
SegFix [8]	65.83	2961	0.38	-	-	-	-	-	-
MagNet-Fast	66.91	2007	0.32	71.85	1559	0.29	69.75	2621	0.33
MagNet	**67.57**	2007	2.93	**72.96**	1559	1.19	**70.60**	2621	2.74



Conclusion

❑ Introduce a novel **semantic segmentation framework** for high-resolution images without exploding GPU memory.

❑ With patches of multiple scales, predictions are refined progressively from the coarsest to the finest levels.

❑ MagNet outperforms the previous SOTA methods by a margin of 1- 2% in mIoU.

Related works

	DenseCRF [1]	GF [2]	DGF [3]	ISS [4]	GLNet [5]	CascadePSP [6]	PointRend [7]	SegFix [8]	DeepStrip [9]	Ours
Deep learning			✅	✅	✅	✅	✅	✅	✅	✅
High-res input	✅	✅	✅		✅	✅		✅	✅	✅
Multi-scale processing					✅	✅	✅			✅
New objects recognization	✅	✅	✅	✅	✅	✅	✅			✅
Partly/Locally refinement							✅	✅	✅	✅

1. Krahenbuhl et al., *Efficient inference in fully connected CRFs with Gaussian edge potentials*. In NeurIPS, 2011.
2. He et al., *Guided image filtering*. In ECCV, 2010.
3. Wu et al., *Fast end-to-end trainable guided filter*. In CVPR, 2018.
4. Li et al., *Iterative instance segmentation*. In CVPR, 2016.
5. Chen et al., *Collaborative global-local networks for memory-efficient segmentation of ultra-high resolution images*. In CVPR, 2019.
6. Cheng et al., *CascadePSP: Toward class-agnostic and very high-resolution segmentation via global and local refinement*. In CVPR, 2020.
7. Kirillov et al., *PointRend: Image segmentation as rendering*. In CVPR, 2020.
8. Yuan et al., *SegFix: Model-agnostic boundary refinement for segmentation*. In ECCV, 2020.
9. Zhou et al., *DeepStrip: High-resolution boundary refinement*. In CVPR, 2020.
10. Cordts et al., *The Cityscapes dataset for semantic urban scene understanding*. In CVPR, 2016.
11. Demir et al., DeepGlobe 2018: A challenge to parse the earth through satellite images. 2018.
12. Walker et al., MICCAI automatic prostate gleason grading challenge. 2019.

Code is available at:

https://github.com/VinAIResearch/MagNet