

EDEN: Multimodal Synthetic Dataset of Enclosed garDEN Scenes

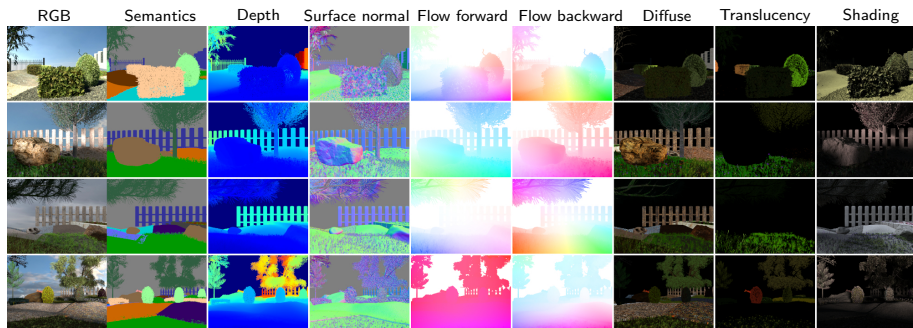
Hoang-An Le¹, Thomas Mensink^{1,2}, Partha Das^{1,3},
Sezer Karaoglu^{1,3}, Theo Gevers^{1,3}

¹Computer Vision Lab, University of Amsterdam,

²Google Research, Amsterdam, ³3D Universum

<https://lhoangan.github.io/eden>

Overview



Outline

1 Dataset

- Modelling
- Rendering

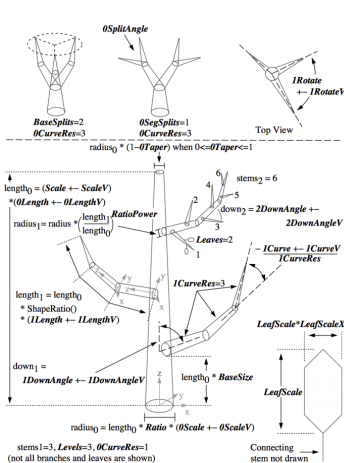
2 Experiment

- Semantic segmentation
- Monocular depth prediction

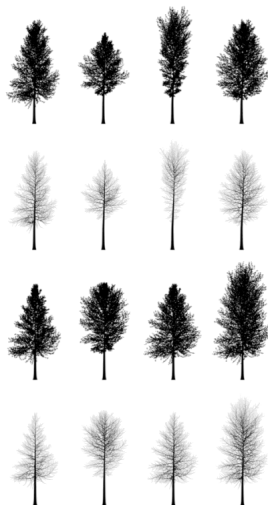
3 Conclusion

Trees

Parametric tree model: (Sapling Blender Addon)

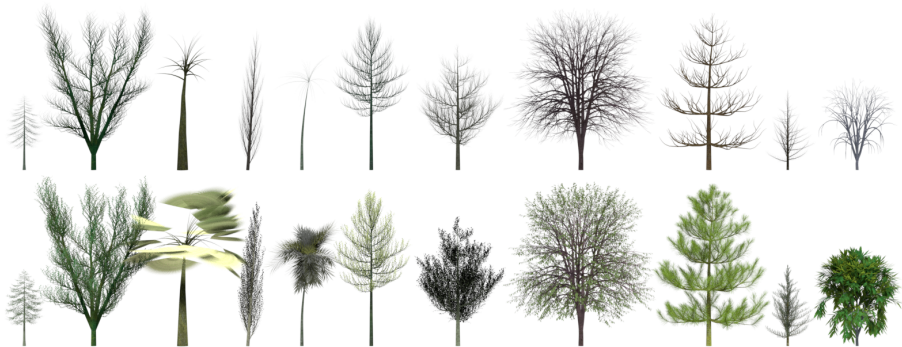


Tree diagram from *Creation and Rendering of Realistic Trees*, Jason Weber and Joseph Penn. ©1995 ACM, Inc.



Trees

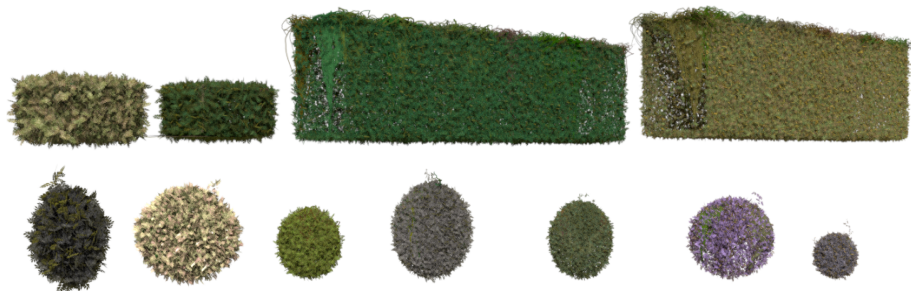
Parametric tree model: (Sapling Blender Addon)



Sample tree models (top: tree stems, bottom: with leaves) for various tree species

Bushes

Parametric Ivy model: (IvyGen Blender Addon)



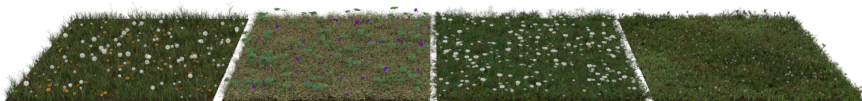
Sample models for hedges (top) and topiaries (bottom). The bushes can be generated with various sizes, leaf colors, and internal stem structures.

Landscapes and terrain

Grass models: (The Grass Essentials package) parametric appearances



15 grass and 15 weed species, up to 49 model variations for each



Landscapes and terrain

Road textures: ([Poliigon](#)) each comprise professionally designed intrinsic components: albedo, normal, glossy, etc.

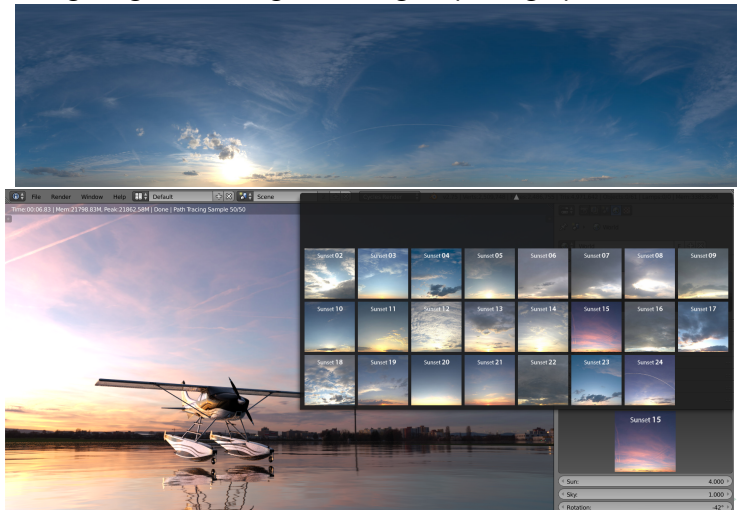


Sample tiles of different terrains: gravel, pavement, pebble stones, dirt.

Environment

Environmental lighting and sky: (The Pro-Lighting Skies package)

- HDR lighting uses a single 360-degree photograph



Environment

Environmental lighting and sky: (The Pro-Lighting Skies package)

- HDR lighting uses a single 360-degree photograph

clear



cloudy



overcast



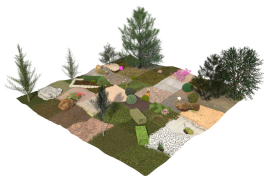
sunset



twilight

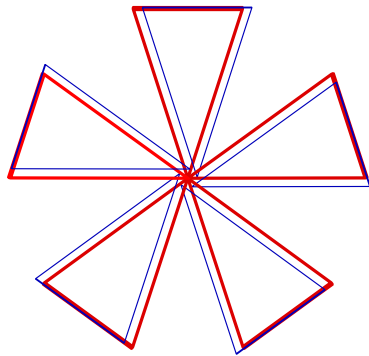
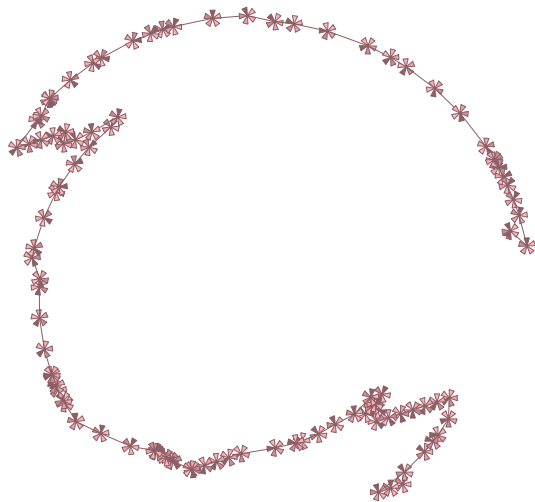


Garden construction



Camera setup

5 pairs of stereoscopic cameras, randomly rotated and moved for 100 steps



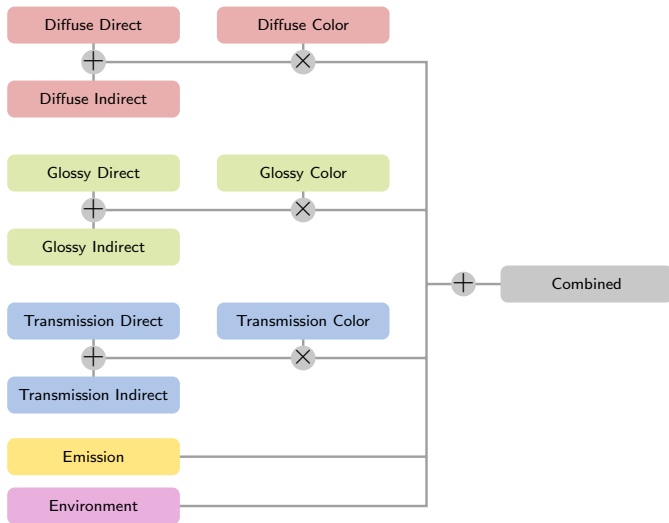
Camera setup

5 pairs of stereoscopic cameras, randomly rotated and moved for 100 steps



Examples of the generated trajectories used in the rendering process. The 5 pairs of cameras, illustrated by different color shades, are randomly moved, turned, and self-rotated while avoiding obstacles in a garden.

Blender Cycles render engine



Outline

- 1 Dataset
 - Modelling
 - Rendering
- 2 Experiment
 - Semantic segmentation
 - Monocular depth prediction
- 3 Conclusion

Semantic segmentation

- DeepLabv3+ architecture with Xception-65 backbone
- Analyzing 3 aspects:
 - training and test size
 - lighting conditions
 - realism and compatibility with real-world dataset
- Label set, following the 3DRMS challenge
 - unknown (0, to be ignored), grass, ground, pavement, hedge, topiary, flower, obstacle, tree, background (sky, and far objects)
 - a small 20K test set is proposed for reducing test time

Training and testing size

Sampling	test	
	full	20K
25%	75.71	75.89
50%	79.42	79.52
100%	81.96	82.09

The network performance increase when being trained on higher number of images

- showing the advantage of having large number of training images

The performance on the reduced test set is on par with the full set

- the reduced test set is used unless mentioned otherwise

Cross-lighting analysis

Training	test					
	clear	cloudy	overcast	sunset	twilight	20K
clear	76.10	<i>76.91</i>	76.43	72.23	75.91	72.03
cloudy	75.09	77.59	77.16	72.37	76.40	72.30
overcast	65.75	75.52	78.41	70.76	74.63	70.22
sunset	73.21	75.76	77.17	74.44	77.28	71.84
twilight	66.19	72.86	76.21	70.55	78.16	68.83

- Row: a model trained on the specific lighting condition (highest in *italics*)
- Column: the results evaluated on the specific subset (highest in **boldface**)
- Lighting-specific training gives better results on the specific lighting
- Cross-lighting vary depending on the conditions of the training and test images.

Real-world datasets

Method

- Pre-training a semantic segmentation network with different dataset
- Freezing encoder + finetuning decoder on target dataset's training split

Baselines

- Generic: pretrained on ImageNet + COCO + PASCAL-VOC 2012
- Cityscapes: pretrained on ImageNet + Cityscape

Target datasets

- 3DRMS: garden scenes, 221 training, 228 test, 10 classes
- Freiburg: forest scenes, 228 training, 15 test, 6 classes

Pre-training	test	
	3DRMS	Freiburg
Generic	24.35	41.33
Cityscapes	31.11	50.08
EDEN	34.55	52.45

Adaptability of pre-trained features to unstructured natural real-world scenes

Monocular depth prediction

Performance of SOTA methods for monocular depth prediction on KITTI and EDEN

- supervised methods outperform unsupervised methods
- the gap among EDEN is larger than among KITTI

Method	Supervised	Dataset	rel	log10	rms
MD2	None	KITTI	0.115	0.193	4.863
VNL	Depth	KITTI	0.072	0.117	3.258
MD2	None	EDEN	0.438	0.556	1.403
MD2	Pose	EDEN	0.182	0.220	0.961
VNL	Depth	EDEN	0.181	0.083	1.061

Outline

- 1 Dataset
 - Modelling
 - Rendering
- 2 Experiment
 - Semantic segmentation
 - Monocular depth prediction
- 3 Conclusion

Conclusion

EDEN: Large-scale Multimodal Synthetic Dataset of Enclosed garDEN Scenes

- focusing on outdoor unstructured scenes (gardens, parks) and agricultural scenarios
- +300K images with various computer vision modalities
- simulating different lighting condition: clear sky, cloudy, overcast, sunset, twilight
- experiments on semantic segmentation show realistic features for natural scenes.

More information at

<https://lhoangan.github.io/eden>

Thank you for watching