ENGN8501: Advanced Topics in Computer Vision
Reading Group 1

Week 2

Dylan Campbell

dylan.campbell@anu.edu.au

Australian National University

Semester 2, 2020

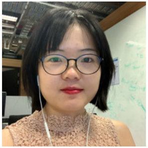
Literature Reading Group: Logistics

Logistics (1)

Your tutors:

Sameera Ramasinghe

Lin Li



Logistics (2)

- **Presenter:** Dylan
- **Paper:** Wang et al., "SDC-Depth: Semantic Divide-and-Conquer Network for Monocular Depth Estimation", CVPR, IEEE/CVF, 2020
- **You will have**
  1. read the paper; and
  2. prepared two questions about it;
- and are expected to **participate** in the discussion.
  - Verbally or via chat
  - Ask questions, answer questions, make relevant comments

Logistics (3)

- **Presenter schedule:** On Wattle now. Please check to see when you have been scheduled to present.
- **Zoom name:** Make sure your name is set to your **full name** and **student ID**; click on Participants → More (next to your name) → Rename
- **Breakout rooms:** After the presentation, we will divide into 3 breakout rooms and the tutors or I will facilitate the discussion.

Run sheet (Week 3 onwards)

1. Divide into 3 breakout rooms
2. **Paper A:** 10–15 minute summary presentation
3. **Paper A:** 15 minute discussion
4. **Paper B:** 10–15 minute summary presentation
5. **Paper B:** 15 minute discussion

FAQs (1)

- How detailed should the presentation be?
  - It's a summary or overview, it does not need to be too detailed.
  - Cover the problem, motivation, main ideas, and results.
  - Copying formulas/diagrams directly from the paper is allowed.
- What should the discussion questions be like?
  - Intended to provoke discussion.
  - Often things that you find confusing or difficult to explain.
  - "Why" or "how" questions are good, such as "Why does batch normalisation improve the stability of training?" or "How does the XYZ module lead to the results presented?"
  - Things that are not explained clearly or are assumed in the paper might be good candidates.

FAQs (2)

- What if no one responds to my discussion questions?
  - The tutor (or I) will be there to facilitate.
  - It will not affect your mark if no one responds.
  - There are marks for participation, so if no one responds, then they may do less well for this assessment item.
  - It may be that no-one in your group is able to contribute something to a particular discussion question—this is okay.
  - Either the tutor will help, or you can note down the question and post it to the Wattle forums.
- Do we need to know the answer for our questions? No.
- If I am presenting and cannot answer a question, will this affect my mark? No.

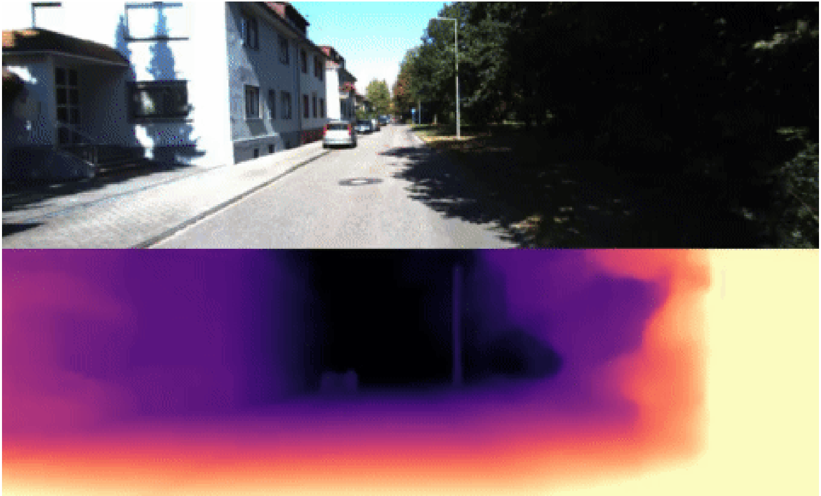
SDC-Depth: Semantic Divide-and-Conquer Network for Monocular Depth Estimation

Lijun Wang, Jianming Zhang, Oliver Wang, Zhe Lin, and Huchuan Lu

CVPR 2020

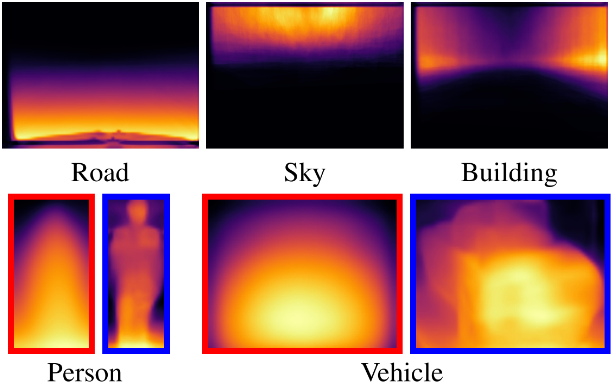
Monocular depth estimation

The Problem: Compute per-pixel depth from a single photo.



Credit: Godard et al., “Digging into Self-Supervised Monocular Depth Prediction”, ICCV 2019

Key insights



- Semantic categories have consistent depth structures
- Relative depth maps of different instances are similar
- Absolute depth of an instance depends on its relative size and position in scene

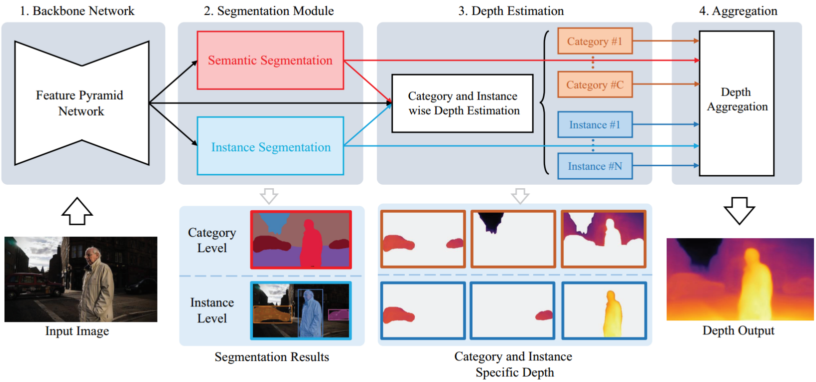
Credit: Wang et al., "SDC-Depth: Semantic Divide-and-Conquer Network for Monocular Depth Estimation", CVPR 2020

Contributions

The authors list their contributions as

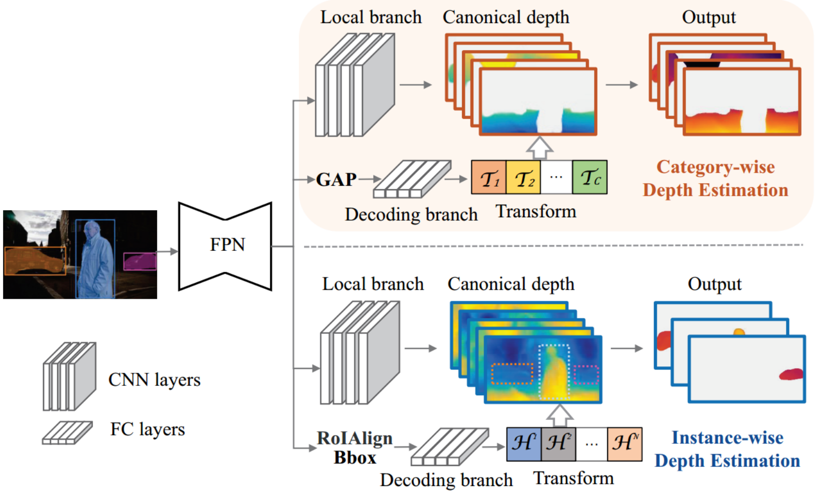
1. a novel framework for monocular depth estimation based on a semantic *divide-and-conquer* strategy;
2. ~~an implementation~~ [not a contribution]; and
3. ~~state-of-the-art results~~ [not a contribution].

Semantic divide-and-conquer network



1. Decompose image via semantic and instance segmentation
2. Estimate normalised depth per segment + transformation
3. Aggregate depth estimates

Per-segment depth estimation



- Separate category and instance streams:
  - Local branch: convolution layers for depth
  - Decoding branch: scale and shift transformations

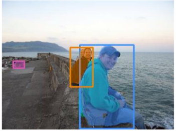
Depth aggregation

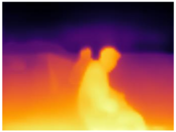
1. Local updates to category depth maps using instance depths
2. Aggregate depth maps using semantic segmentation

- Aggregation is differentiable, allowing end-to-end training
- Multi-task training with semantic segmentation (L_{S}), instance segmentation (L_{I}), and depth (L_{D}) losses:
  $L = \alpha_{\mathrm{S}} L_{\mathrm{S}} + \alpha_{\mathrm{I}} L_{\mathrm{I}} + \alpha_{\mathrm{D}} L_{\mathrm{D}}$


Input


Semantic Segmentation

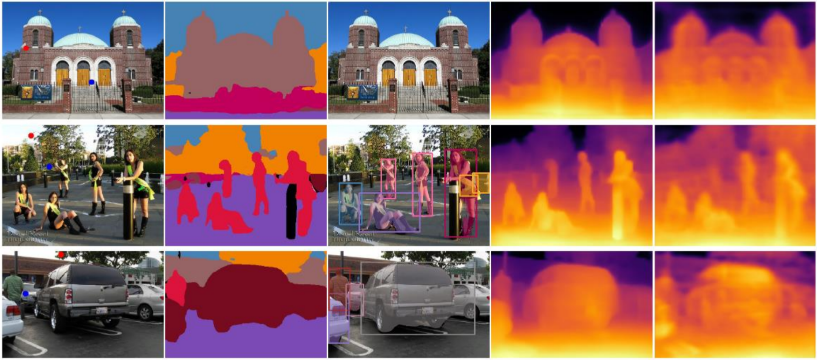

Instance Segmentation


Depth Estimation

Evaluation on Depth in the Wild

- Depth trained using annotations consisting of sparse ordinal pointwise information

Method	Chen *et al.* [2]	Xian *et al.* [37]	Xu *et al.* [39]	Ours
WHDR	22.14%	14.98%	13.02%	**11.21%**



Input   Our Segmentation   Our Depth   Baseline Depth

Evaluation on Cityscapes

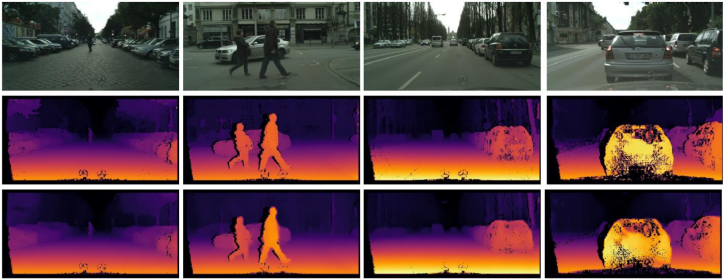
- Depth trained using dense depth annotations

Method	Error				Accuracy		
	RMSE	RMSE (log)	Abs Rel	Sq Rel	$\delta < 1.25$	$\delta < 1.25^2$	$\delta < 1.25^3$
Laina *et al.* [14]	7.273	0.448	0.257	4.238	0.765	0.893	0.940
Xu *et al.* [39]	7.117	0.428	0.246	4.060	0.786	0.905	0.945
Zhang *et al.* [41]	7.104	0.416	0.234	**3.776**	0.776	0.903	0.949
Ours	**6.917**	**0.414**	**0.227**	3.800	**0.801**	**0.913**	**0.950**

Input

GT

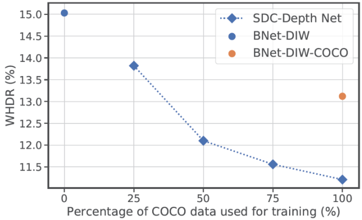
Ours



Ablation study

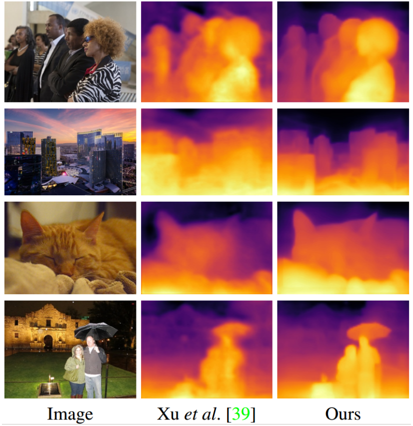
- Which source of information is most helpful?

		SDC-A	SDC-B	SDC-C	SDC-D
Design Choice	Cat.	✗	✓	✓	✓
	Ins.	✗	✗	✓	✓
	DEnt.	✗	✗	✗	✓
Err.	RMSE	7.203	6.962	6.958	**6.917**
	Abs Rel	0.276	0.236	0.234	**0.227**
Acc.	$\delta < 1.25$	0.767	0.794	0.797	**0.801**
	$\delta < 1.25^2$	0.895	0.911	0.911	**0.913**
	$\delta < 1.25^3$	0.941	0.949	0.951	**0.950**



Summary

- A divide-and-conquer strategy for per-segment depth estimation
- Aggregation strategy that makes full use of category and instance-level information
- Impressive results



Discussion

[To breakout rooms]

- Why is object instance information important?
- Why are affine transformations learned and what are the limitations?
- Why are the instance depth maps weighted by p^i?
- What are the limitations of the method in general?

Resources:

- Hui, “Understanding Feature Pyramid Networks for object detection (FPN)”, Medium, 2018
- Tsang, “Review: FCN — Fully Convolutional Network (Semantic Segmentation)”, Medium, 2018
- Odena et al., “Deconvolution and Checkerboard Artifacts”, Distill, 2016