Panoptic Segmentation
J. Yuan
Introduction
PanUrban Dataset
Dataset Annotation
Dataset Features & Statistics
Panoptic Segmentation
Entry
(R)RPN
(R)ROI
Semantic
Experiments
Metrics
Pretraining
Augmentation
Box fromat comparasion
Cross Inference
Conclusions
References
References
appendix

Panoptic Segmentation on aerial Images in urban Area

Jie Yuan
Matrikel-Nr.: 10016146
Supervisor: Prof. Christian Heipke
M.Sc. Lin Chen

Institut für Photogrammetrie und GeoInformation
Leibniz Universität Hannover

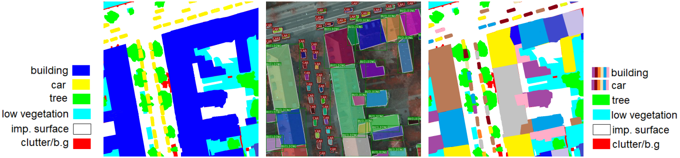
March 6, 2021

Table of contents

1. Introduction
2. PanUrban Dataset Annotation
   - Annotation Workflow
   - Dataset Features & Statistics
3. Panoptic Segmentation Model
   - Entry Net
   - Region Proposal Network
   - Object Classification & Mask Prediction
   - Semantic Branch
4. Experiments
   - Task Metrics
   - Pretraining
   - Full Range Augmentation
   - ABB vs. RBB
   - Cross Dataset Performance
5. Conclusions

Panoptic Segmentation

What is Panoptic Segmentation?



semantic　　instance　　panoptic

- countatble things vs. uncountable stuff.
- panoptic - including everything visible in one view.
- a unified framework combining instance segmentation & semantic segmentation.
- applications on automatic building map updating.
- lacking of datasets on aerial images.

1 Introduction

2 PanUrban Dataset Annotation
- Annotation Workflow
- Dataset Features & Statistics

3 Panoptic Segmentation Model
- Entry Net
- Region Proposal Network
- Object Classification & Mask Prediction
- Semantic Branch

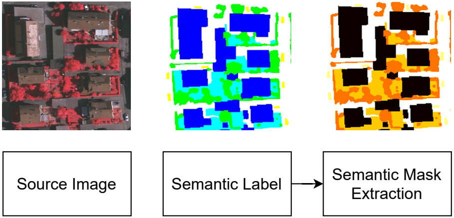
4 Experiments
- Task Metrics
- Pretraining
- Full Range Augmentation
- ABB vs. RBB
- Cross Dataset Performance

5 Conclusions

Introduction
PanUrban Dataset
Dataset Annotation
Dataset Features & Statistics
Panoptic Segmentation
Entry
(R)RPN
(R)ROI
Semantic
Experiments
Metrics
Pretraining
Augmentation
Box fromat comparasion
Cross Inference
Conclusions
References
References
appendix

Dataset Annotation

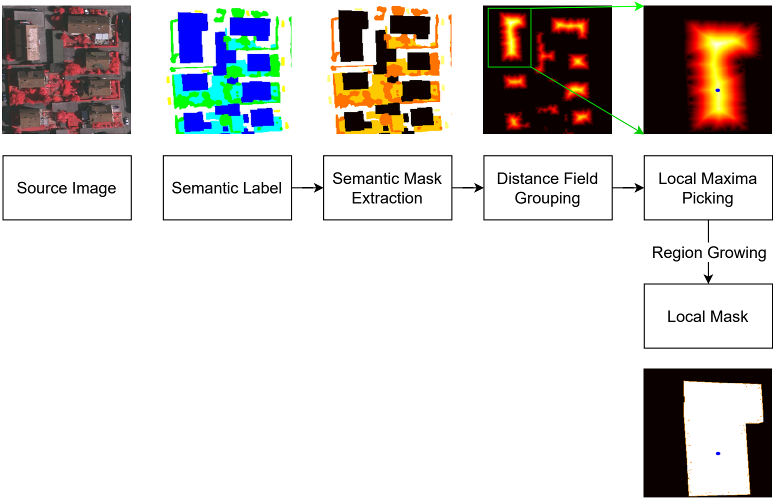
From Semantic to Instance



Introduction
PanUrban Dataset
Dataset Annotation
Dataset Features & Statistics
Panoptic Segmentation
Entry
(R)RPN
(R)ROI
Semantic
Experiments
Metrics
Pretraining
Augmentation
Box fromat comparasion
Cross Inference
Conclusions
References
References
appendix

Dataset Annotation

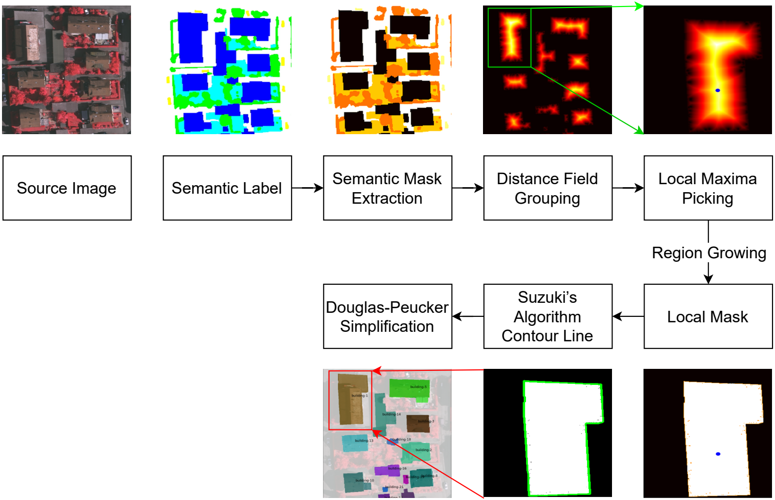
From Semantic to Instance



Introduction
PanUrban Dataset
Dataset Annotation
Dataset Features & Statistics
Panoptic Segmentation
Entry
(R)RPN
(R)ROI
Semantic
Experiments
Metrics
Pretraining
Augmentation
Box fromat comparasion
Cross Inference
Conclusions
References
References
appendix

Dataset Annotation

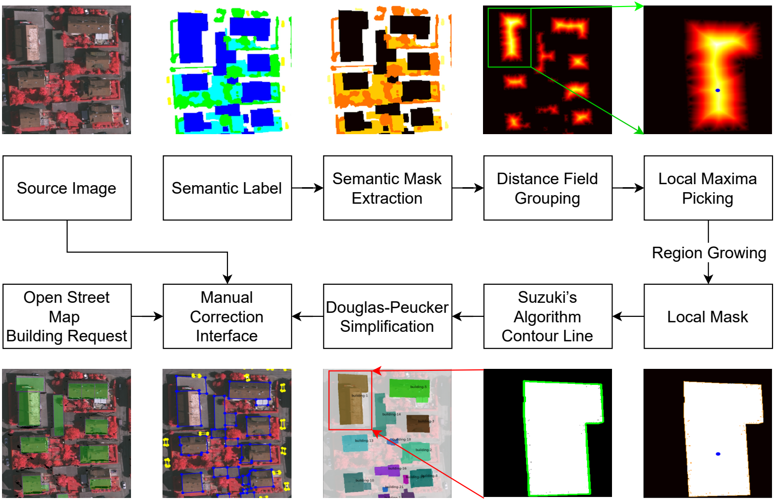
From Semantic to Instance



Introduction
PanUrban Dataset
Dataset Annotation
Dataset Features & Statistics
Panoptic Segmentation
Entry
(R)RPN
(R)ROI
Semantic
Experiments
Metrics
Pretraining
Augmentation
Box fromat comparasion
Cross Inference
Conclusions
References
References
appendix

Dataset Annotation

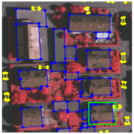
From Semantic to Instance



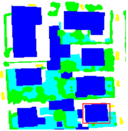
Introduction
PanUrban Dataset
Dataset Annotation
Dataset Features & Statistics
Panoptic Segmentation
Entry
(R)RPN
(R)ROI
Semantic
Experiments
Metrics
Pretraining
Augmentation
Box fromat comparasion
Cross Inference
Conclusions
References
References
appendix

Dataset Annotation

From Instance/Semantic to Panoptic



Instance Annotation

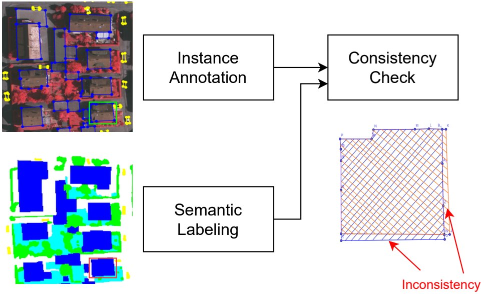


Semantic Labeling

Introduction
PanUrban Dataset
Dataset Annotation
Dataset Features & Statistics
Panoptic Segmentation
Entry
(R)RPN
(R)ROI
Semantic
Experiments
Metrics
Pretraining
Augmentation
Box fromat comparasion
Cross Inference
Conclusions
References
References
appendix

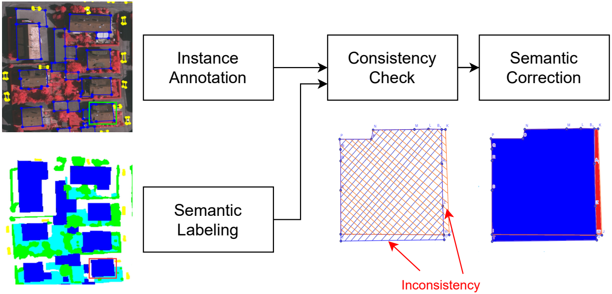
Dataset Annotation

From Instance/Semantic to Panoptic



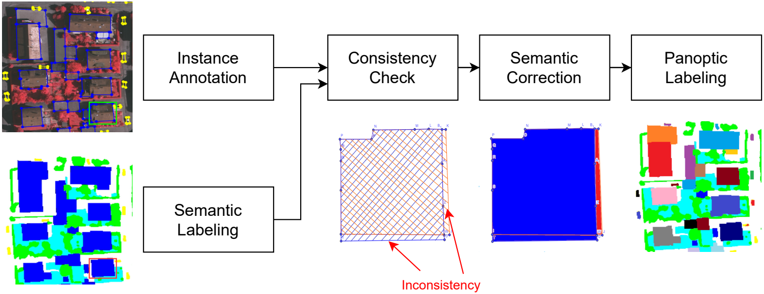
Dataset Annotation

From Instance/Semantic to Panoptic



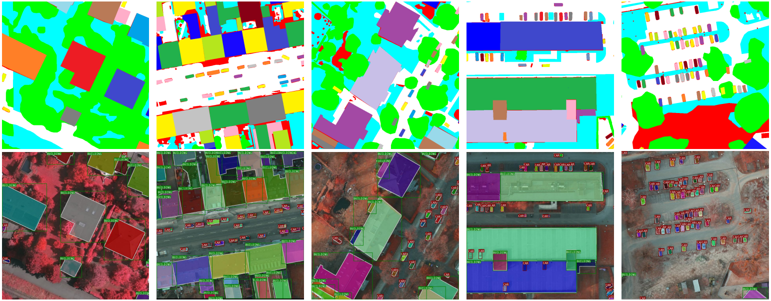
Dataset Annotation

From Instance/Semantic to Panoptic



PanUrban Dataset

Introduction



apartment innercity residual factory parking

- Subsets: Potsdam, Vaihingen, MA.
- Bounding boxes axis-aligned(ABB) or rotated(RBB).
- Compatibility with multiple tasks

PanUrban Dataset

Statistics

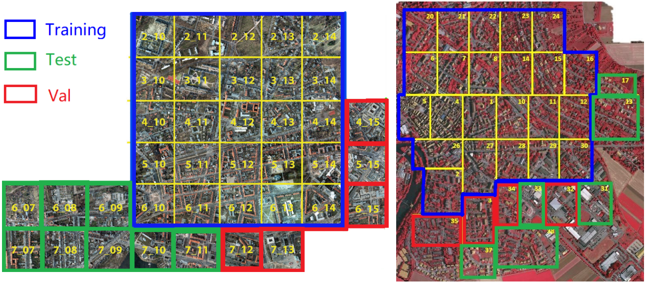
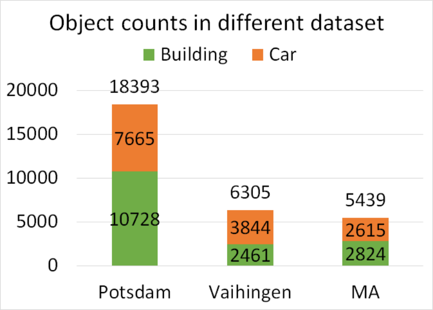


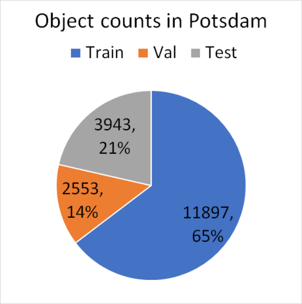
Figure: Dataset split in Potsdam and Vaihingen

- Training is a stitched patch.
- Test and Val are in small image patch.

PanUrban Dataset

Statistics



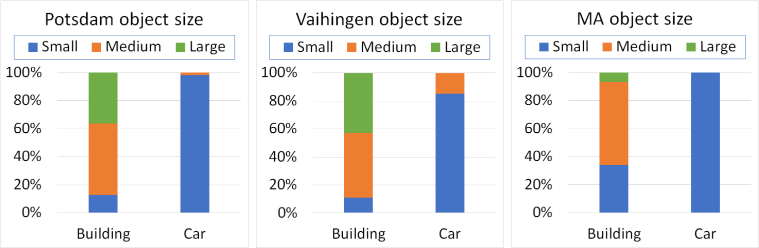


- Number of objects in datasets: Potsdam > Vaihingen > MA
- Number of objects in Train/Val/Test subsets: Train(65%), Test(21%), Val(14%)

Introduction
PanUrban Dataset
Dataset Annotation
Dataset Features & Statistics
Panoptic Segmentation
Entry
(R)RPN
(R)ROI
Semantic
Experiments
Metrics
Pretraining
Augmentation
Box fromat comparasion
Cross Inference
Conclusions
References
References
appendix

PanUrban Dataset

Statistics



Object size s in pixel:

small:$s \in [0, 32^2)$, medium:$s \in [32^2, 96^2)$, large:$s \in [96^2, \infty)$

- most of buildings belongs to medium and large object
- most of cars belongs to small object.

1 Introduction

2 PanUrban Dataset Annotation
- Annotation Workflow
- Dataset Features & Statistics

3 Panoptic Segmentation Model
- Entry Net
- Region Proposal Network
- Object Classification & Mask Prediction
- Semantic Branch

4 Experiments
- Task Metrics
- Pretraining
- Full Range Augmentation
- ABB vs. RBB
- Cross Dataset Performance

5 Conclusions

Model Division

The model of panoptic segmentation can be divided into 6 parts listed as follows:

- **entry network**
- **region proposal network**
- **instance classification**
- **instance mask segmentation**
- **semantic segmentation**
- **fusion**

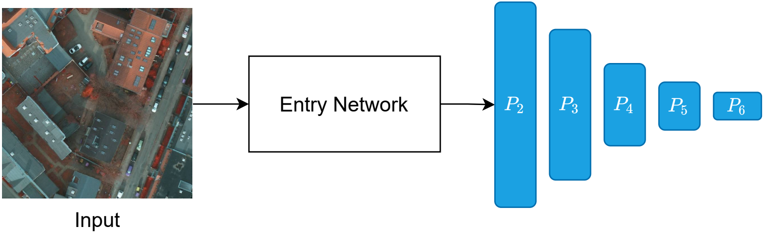
Loss functions:

$$\mathcal{L} = \mathcal{L}_{proposal} + \mathcal{L}_{instance} + \mathcal{L}_{semantic}$$

Introduction
PanUrban Dataset
Dataset Annotation
Dataset Features & Statistics
Panoptic Segmentation
Entry
(R)RPN
(R)ROI
Semantic
Experiments
Metrics
Pretraining
Augmentation
Box fromat comparasion
Cross Inference
Conclusions
References
References
appendix

Entry Net

Entry Net Structure



- input: source image
- feature maps P_i with $i \in [2, 6]$ in different scale.

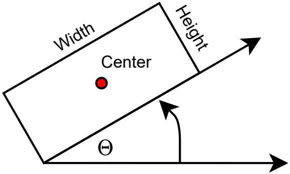
Introduction
PanUrban Dataset
Dataset Annotation
Dataset Features & Statistics
Panoptic Segmentation
Entry
(R)RPN
(R)ROI
Semantic
Experiments
Metrics
Pretraining
Augmentation
Box fromat comparasion
Cross Inference
Conclusions
References
References
appendix

Region Candidates Proposal

Region Proposal Network(RPN)

What is a region? - a bounding box enclosed area.

ABB vs.RBB



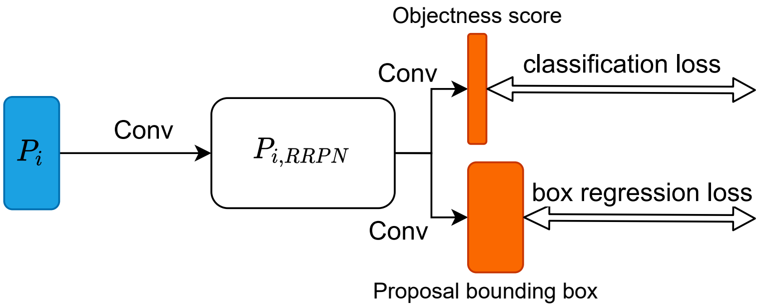
parameterization of RBB

- axis-aligned bounding box(ABB) vs. rotated bounding box(RBB)
- parameterization: (x_c, y_c, w, h).vs. (x_c, y_c, w, h, θ)
- compact alignment with rotated bounding box.

Introduction
PanUrban Dataset
Dataset Annotation
Dataset Features & Statistics
Panoptic Segmentation
Entry
(R)RPN
(R)ROI
Semantic
Experiments
Metrics
Pretraining
Augmentation
Box fromat comparasion
Cross Inference
Conclusions
References
References
appendix

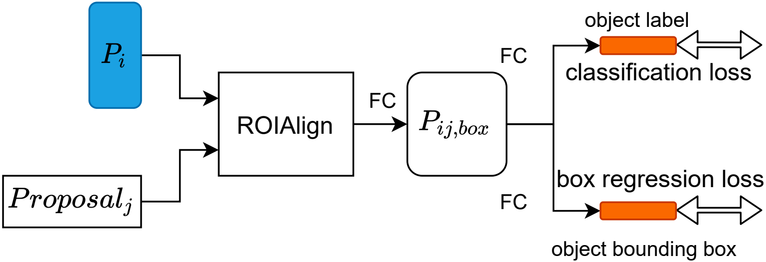
Region Candidates Proposal

Region Proposal Network(RPN) Structure



- Input: feature map from entry network P_i
- Output: objectness score & object bounds
- Loss: $\mathcal{L}_{proposal}^{cls}$ and $\mathcal{L}_{proposal}^{box}$

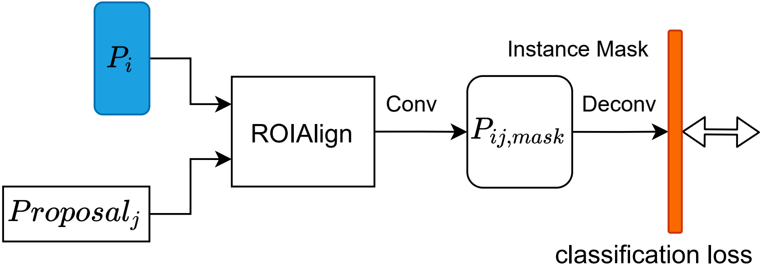
Object Classification



- Input: P_i, $Proposal_j$
- Output: classified object region
- Loss: $\mathcal{L}^{box}_{instance}$ and $\mathcal{L}^{cls}_{instance}$

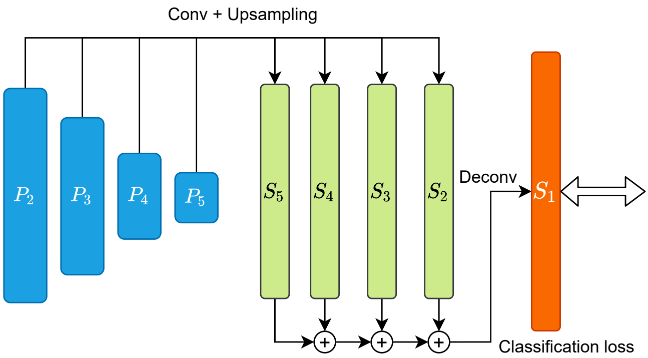
Introduction
PanUrban Dataset
Dataset Annotation
Dataset Features & Statistics
Panoptic Segmentation
Entry
(R)RPN
(R)ROI
Semantic
Experiments
Metrics
Pretraining
Augmentation
Box fromat comparasion
Cross Inference
Conclusions
References
References
appendix

Mask Segmentation



- Input: P_i, $Proposal_j$.
- Output: classified object region
- Loss: $\mathcal{L}_{instance}^{mask}$

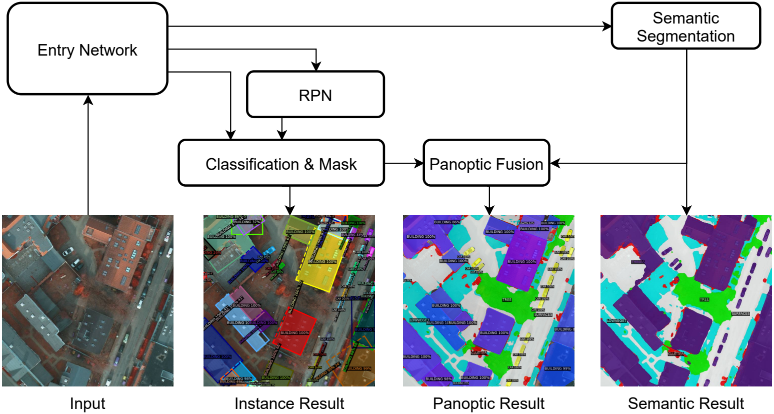
Semantic Segmentation



- input: Multi-scale feature map $P_2 - P_5$
- output: predicted semantic label
- loss: $\mathcal{L}_{semantic}$

Panoptic Segmentation

Full model



1 Introduction

2 PanUrban Dataset Annotation
- Annotation Workflow
- Dataset Features & Statistics

3 Panoptic Segmentation Model
- Entry Net
- Region Proposal Network
- Object Classification & Mask Prediction
- Semantic Branch

4 Experiments
- Task Metrics
- Pretraining
- Full Range Augmentation
- ABB vs. RBB
- Cross Dataset Performance

5 Conclusions

Task Metrics

Panoptic Quality

$$PQ = \frac{\sum_{p,q \in TP} IoU(p,q)}{|TP| + \frac{1}{2}|FP| + \frac{1}{2}|FN|} = \underbrace{\frac{\sum_{p,q \in TP} IoU(p,q)}{|TP|}}_{\text{segmentation quality(SQ)}} \times \underbrace{\frac{|TP|}{|TP| + \frac{1}{2}|FP| + \frac{1}{2}|FN|}}_{\text{recongnition quality(RQ)}}$$

where:

p, q = prediction and groundtruth of matched **segments.**

TP = true positive segment, which has $IoU(p, q) > 0.5$

$|*|$ = count of **segments**, **not pixels.**

SQ = average IoU of **matched segments**

RQ = F1 score of segments.

Experimental Setting

Content of experiment:

1. influence of pretraining on semantic branch.
2. performance comparison on ABB and RBB.
3. inference cross different datasets.

Additional experiment can be found in the thesis:

1. full range augmentation performance.
2. pretraining on RPN.
3. ablation study.
4. parameter study.

Pretraining on Semantic Branch

Pretrained weight is applied on the entry net.

Pretraining\IoU	mIoU	$IoU_{clutter}$	$IoU_{l.veg}$	$IoU_{imp.surf.}$	IoU_{thing}	IoU_{tree}
$COCO_{weight}$	**60.2**	8.9	68.1	**77.2**	**82.9**	**63.8**
$ImageNet_{weight}$	60.1	**13.6**	**68.6**	75.9	79.9	62.5
None	50.6	4.8	56.6	67.4	62.5	61.5
$COCO^{full}_{weight}$	64.1	22.3	69.4	78.6	84.1	66.0

Table: Semantic Metrics w.r.t Pretraining: mIoU denotes averaged Intersection over Union over all classes, subscript hints class name

Result from Semantic Branch:

- Comparison on $COCO_{weight}$ vs. $ImageNet_{weight}$ and None Pretraining.
- Comparison on semantic categories.
- Comparison with fully trained model, inferring the influence of instance branch.

Introduction
PanUrban Dataset
Dataset Annotation
Dataset Features & Statistics
Panoptic Segmentation
Entry
(R)RPN
(R)ROI
Semantic
Experiments
Metrics
Pretraining
Augmentation
Box fromat comparasion
Cross Inference
Conclusions
References
References
appendix

Full range augmentation

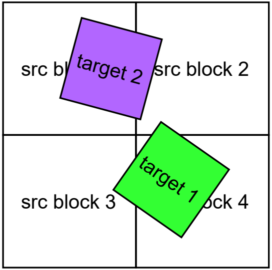


Figure: Example of Full range augmentation on source block 1-4

Features:

- random location
- random orientation
- random small scale difference.
- extract features across source blocks.

Panoptic Segmentation

J. Yuan

Introduction

PanUrban Dataset
Dataset Annotation
Dataset Features & Statistics

Panoptic Segmentation
Entry
(R)RPN
(R)ROI
Semantic

Experiments
Metrics
Pretraining
Augmentation
Box fromat comparasion
Cross Inference

Conclusions

References

References

appendix

ABB vs. RBB

Comparison on bounding box format

Metrics	All classes			Things			Stuff		
	PQ_{all}	SQ_{all}	RQ_{all}	PQ_{th}	SQ_{th}	RQ_{th}	PQ_{st}	SQ_{st}	RQ_{st}
$Potsdam_R$	**54.4**	**74.2**	**72.7**	**70.0**	83.3	**84.2**	**46.6**	**69.6**	**67.0**
$Potsdam_A$	53.9	74.1	72.1	69.4	**84.2**	82.8	46.2	69.1	66.8
$Vaihingen_R$	53.6	63.2	**71.0**	63.9	82.5	77.8	**48.5**	53.6	**67.6**
$Vaihingen_A$	**53.8**	**63.7**	70.8	**64.7**	**82.8**	**78.6**	48.4	**54.1**	67.0

Table: PQ.RQ,SQ in axis-aligned box(ABB) and rotated bounding box(RBB) in dataset Potsdam and Vaihingen, with full range augmentation

- Comparison on Box formats: subscript H and R.
- Comparison on things and stuff.

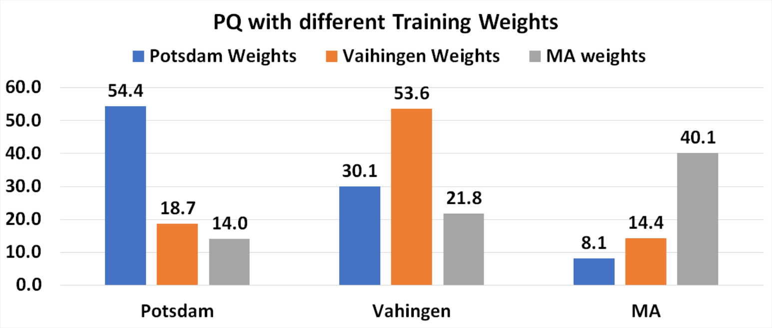
Metircs	PQ_{all}	SQ_{all}	RQ_{all}	PQ_{th}	SQ_{th}	RQ_{th}	PQ_{st}	SQ_{st}	RQ_{st}
MA_R	**40.1**	**58.1**	**56.9**	**47.1**	**74.4**	**62.7**	**36.6**	**49.9**	**54.0**
MA_A	31.5	54.0	48.4	38.3	73.8	51.5	28.1	44.2	46.9

Table: PQ.RQ,SQ on axis-aligned box and rotated box in dataset MA, with rotational augmentation

Introduction
PanUrban Dataset
Dataset Annotation
Dataset Features & Statistics
Panoptic Segmentation
Entry
(R)RPN
(R)ROI
Semantic
Experiments
Metrics
Pretraining
Augmentation
Box fromat comparasion
Cross Inference
Conclusions
References
References
appendix

Cross Dataset Performance

This experiment applies training on a dataset and inference on other datasets.



- Model tend to overfit the training data.

1 Introduction

2 PanUrban Dataset Annotation
- Annotation Workflow
- Dataset Features & Statistics

3 Panoptic Segmentation Model
- Entry Net
- Region Proposal Network
- Object Classification & Mask Prediction
- Semantic Branch

4 Experiments
- Task Metrics
- Pretraining
- Full Range Augmentation
- ABB vs. RBB
- Cross Dataset Performance

5 Conclusions

Conclusions and Outlook

Conclusions:

- PanUrban dataset is developed with a semi-automatic workflow, which can be applied in multiple aerial image tasks.
- Training a panoptic segmentation model has achieved a plausible result which can be set as a baseline in future development.

Future work:

- Limited generalization ability requires techniques to limit overfitting on training data, applying different normalization layers or add regularization terms.

References I

- Ross Girshick. “Fast r-cnn”. In: *Proceedings of the IEEE international conference on computer vision*. 2015, pp. 1440–1448.
- Kaiming He et al. “Mask r-cnn”. In: *Proceedings of the IEEE international conference on computer vision*. 2017, pp. 2961–2969.
- Alexander Kirillov et al. “Panoptic feature pyramid networks”. In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 2019, pp. 6399–6408.
- Alexander Kirillov et al. “Panoptic segmentation”. In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 2019, pp. 9404–9413.

Introduction

PanUrban Dataset

Dataset Annotation

Dataset Features & Statistics

Panoptic Segmentation

Entry

(R)RPN

(R)ROI

Semantic

Experiments

Metrics

Pretraining

Augmentation

Box fromat comparasion

Cross Inference

Conclusions

References

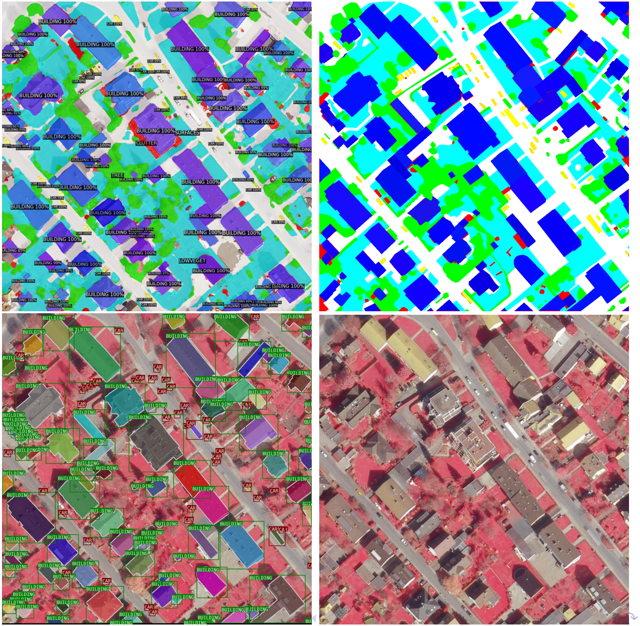
References

appendix

References II

Shaoqing Ren et al. "Faster r-cnn: Towards real-time object detection with region proposal networks". In: *arXiv preprint arXiv:1506.01497* (2015).

Prediction Samples - MA



Prediction Samples - Potsdam