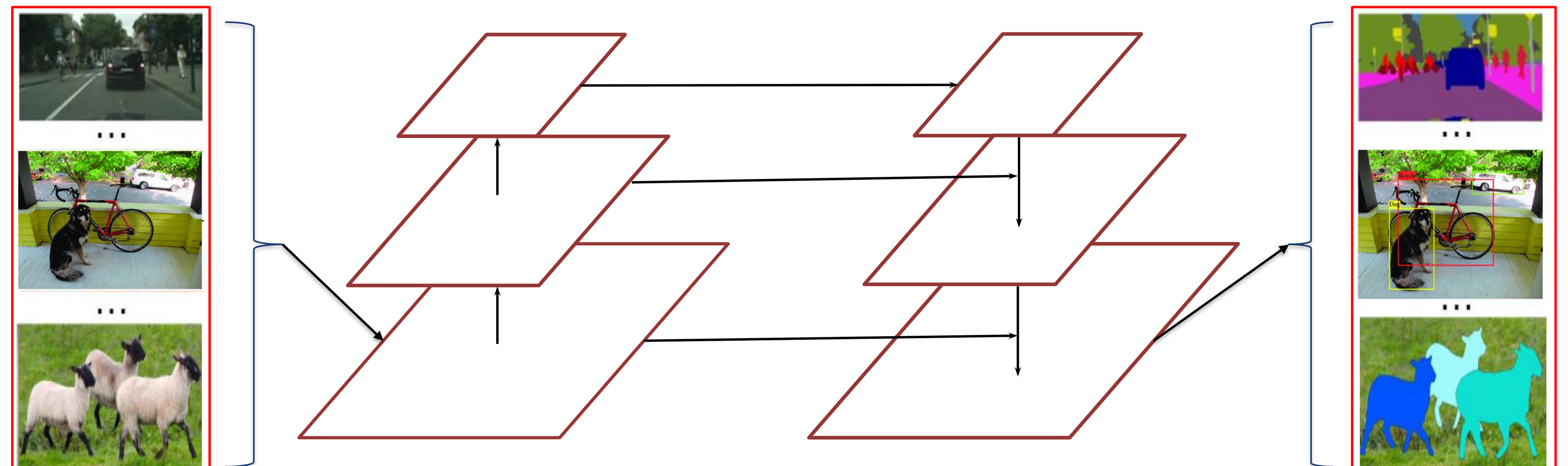
FaPN: Feature-aligned Pyramid Network for Dense Image Prediction

Shihua Huang, Zhichao Lu, Ran Cheng, and Cheng He
Southern University of Science and Technology

Dense Image Prediction:



Feature pyramid network for dense image prediction

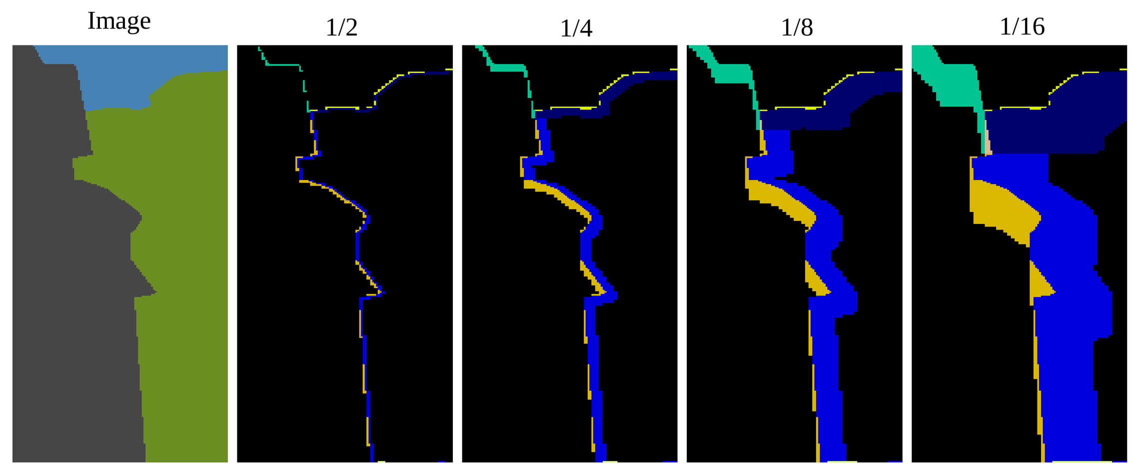
Feature pyramid network (FPN) is the most popular framework for dense image prediction which contains two parts:

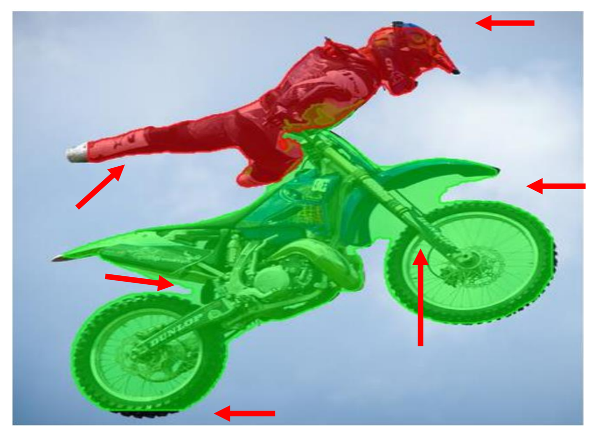
- The left is used to reduce the resolution of features for strong semantics.
- The right aims to backward propagate extracted semantics to each scale.

Motivation:

- Step-by-step downsampling makes the features achieve strong semantic while losing details progressively and dramatically.
- Upsampling the coarse feature without any location reference would misplace semantics into wrong positions, i.e., misaligned context.
- The locality of convolutional and upsampling operations lead to the local scope of misaligned context in which object boundaries suffer from severe misclassification due to the ambiguous context from the nearby different objects.

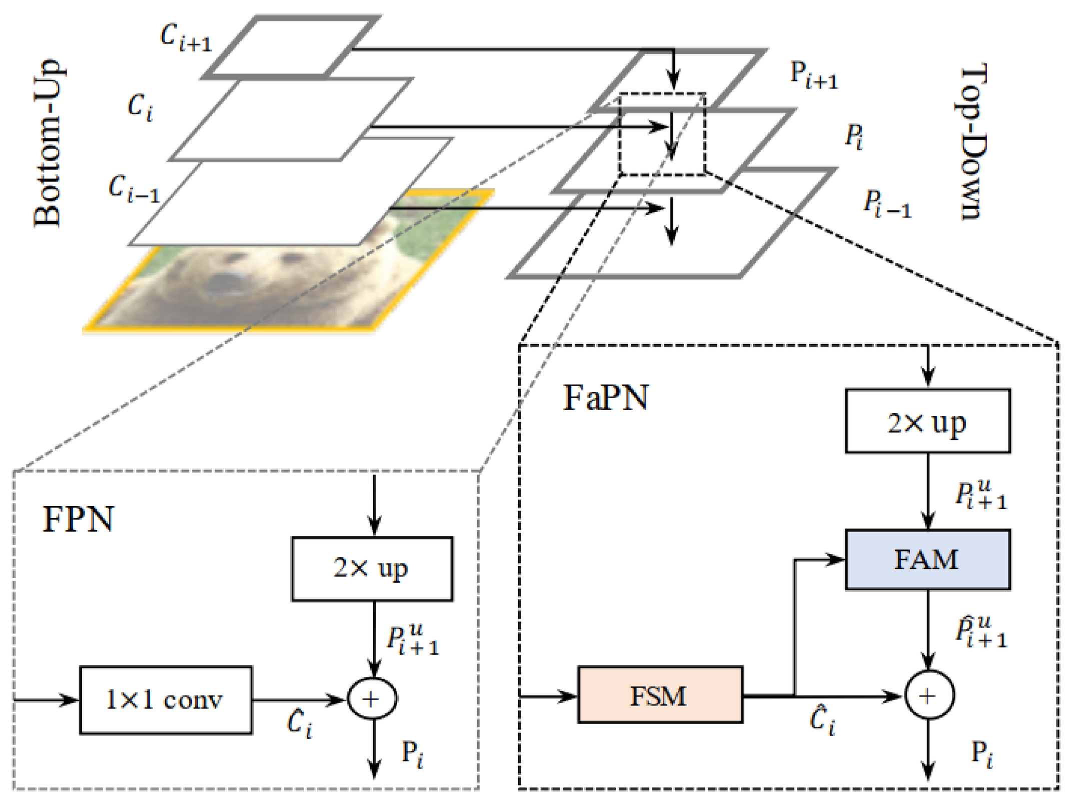
Two examples to illustrate the misaligned boundaries:



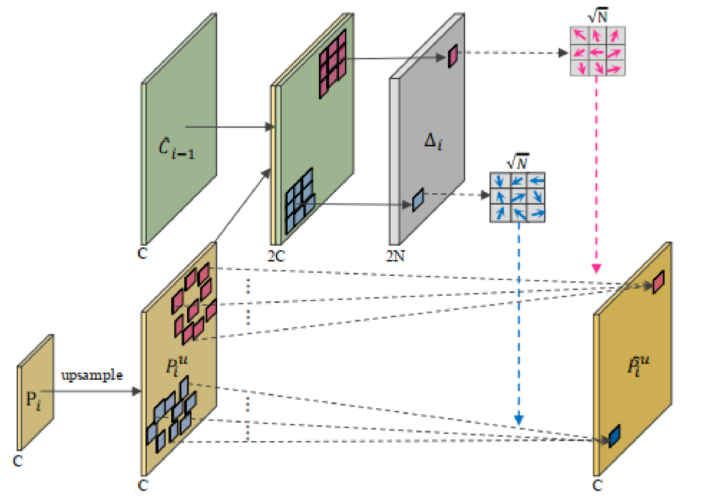


Differences between the image and image after rescaled (left); An example output from Mask-RCNN (right).

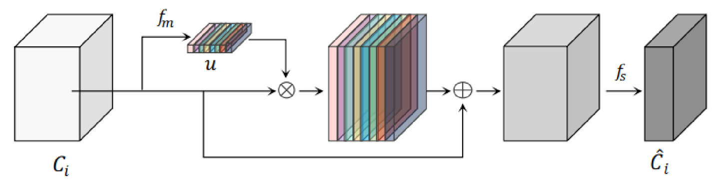
Feature-aligned Pyramid Network:



General comparison between FPN and FaPN



FAM: Feature alignment module



FSM: Feature selection module

- Compared to FPN, the proposed FaPN has two additional modules.
- FaPN is flexible and can be placed in any FPN-based framework.
- FAM firstly learns the offsets from the differences between the detailed and upsampled features and then aligns the upsampled features with the learned offsets.
- FSM aims to model the importance of each feature map in detailed features and emphases the rich detailed features by multiplying the importance values before channel reduction.

Ablation Study:

method	backbone	#Params (M)	mIoU (%)
FPN	R50	28.6 (+4.5)	77.4 (+2.6)
FPN + extra 3×3 conv.	R50	33.4 (-0.3)	77.5 (+2.5)
FPN	R101	47.6 (-14.5)	78.9 (+1.1)
FPN + FAM	R50	31.7 (+1.4)	79.7 (+0.3)
FPN + FAM + SE	R50	33.1 (+0.0)	78.8 (+1.2)
FPN + FAM + FSM (FaPN)	R50	33.1 (+0.0)	**80.0** (+0.0)
FPN + deconv + FSM	R50	32.7 (+0.4)	76.7 (+3.3)
FPN + FAM† + FSM	R50	32.7 (+0.4)	79.3 (+0.7)

- # 2~3: More learnable parameters in FPN brings limited improvement.
- # 4~5: FAM is compatible with FSM, while SE harms the performance.
- # 7: Learnable upsampling could not address feature alignment.
- # 8: Location reference matters.

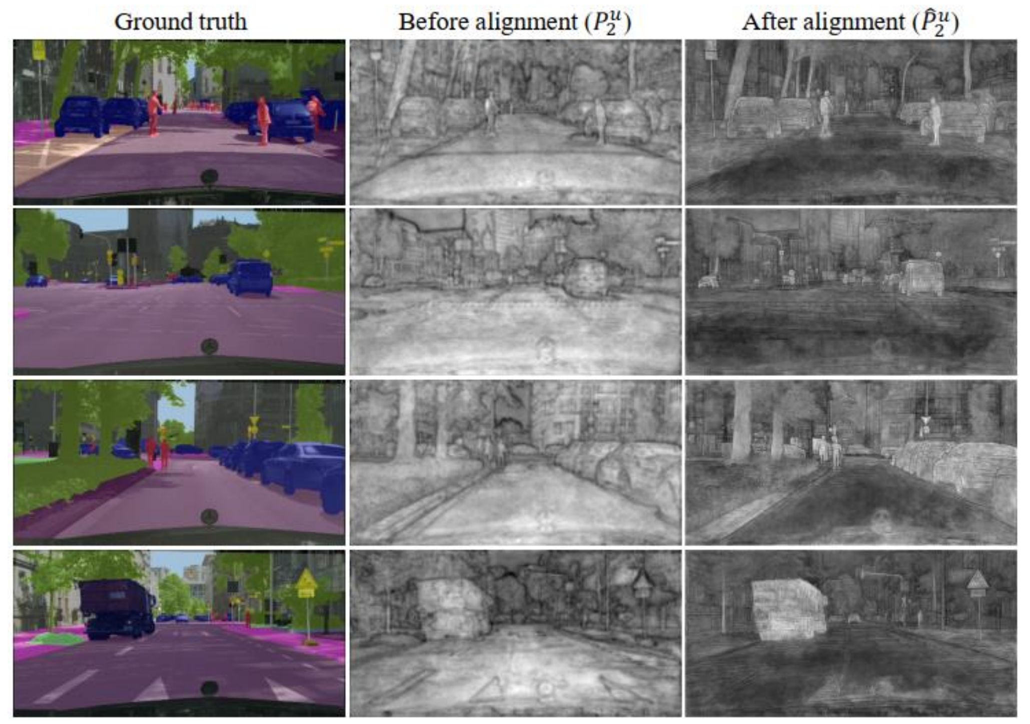
Boundary Prediction Analysis:

method	backbone	3px	5px	8px	12px	mean
FPN	PointRend [21] R50	46.9	53.6	59.3	63.8	55.9
FaPN		49.2	56.2	62.0	66.4	58.5
improvement		(+2.3)	(+2.6)	(+2.7)	(+2.6)	(+2.6)
FPN	PointRend [21] R101	47.8	54.6	60.5	64.9	57.0
FaPN		50.1	57.1	62.9	67.2	59.3
improvement		(+2.3)	(+2.5)	(+2.4)	(+2.3)	(+2.3)

Segmentation performance around boundaries

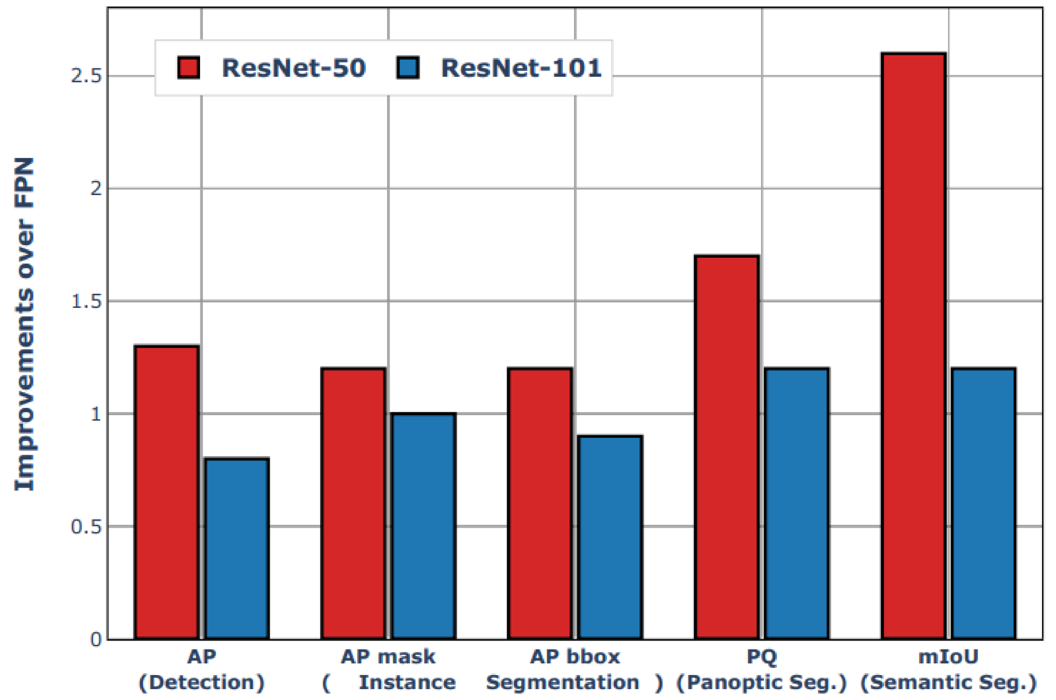
Both the quantitative evaluation and qualitative observation are consistent:

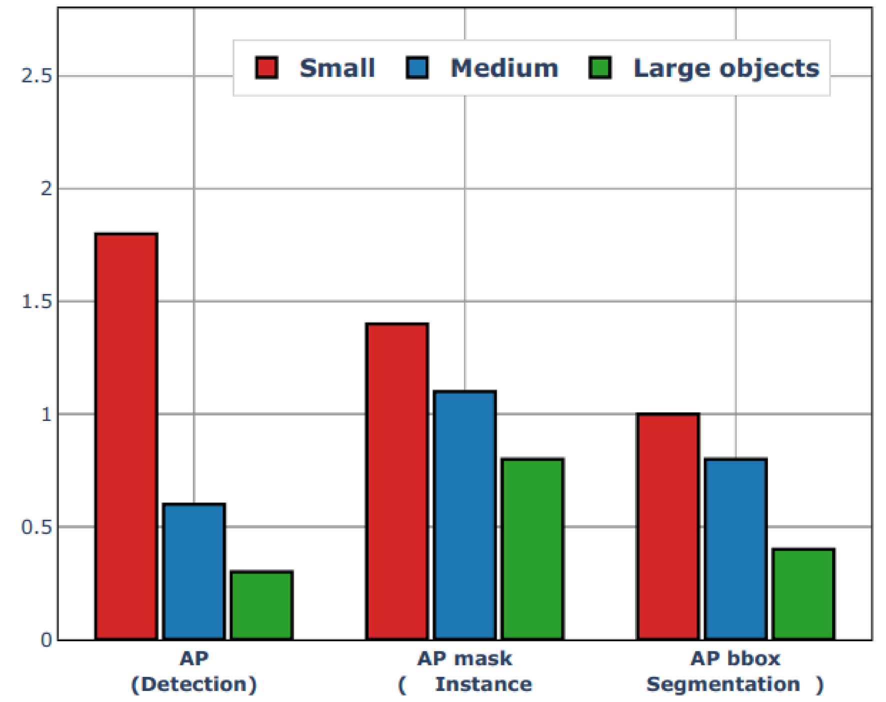
- FaPN achieves higher mIoU on the boundary segmentation.
- Aligned features are smooth and containing more precise object boundaries.



Visualization of the input to and the output from FAM

Main Results:





- Our FaPN can be applied in four dense image prediction tasks.
- A simple replacement of FPN with FaPN in five representative methods yields an overall improvement of 1.2 - 2.6 points in AP / mIoU.
- Our FaPN mainly boosts the performance of small objects.

Further explorations:

- When integrated within MaskFormer, FaPN achieves 56.7% mIoU on ADE20k.
- FaPN can be easily extended to real-time semantic segmentation by pairing it with a ResNet18 which obtains competitive results against dedicated methods.

The code is available at https://github.com/ShihuaHuang95/FaPN-full.