**Computer Vision Foundations – Exercise 5 – Zhao Sun**

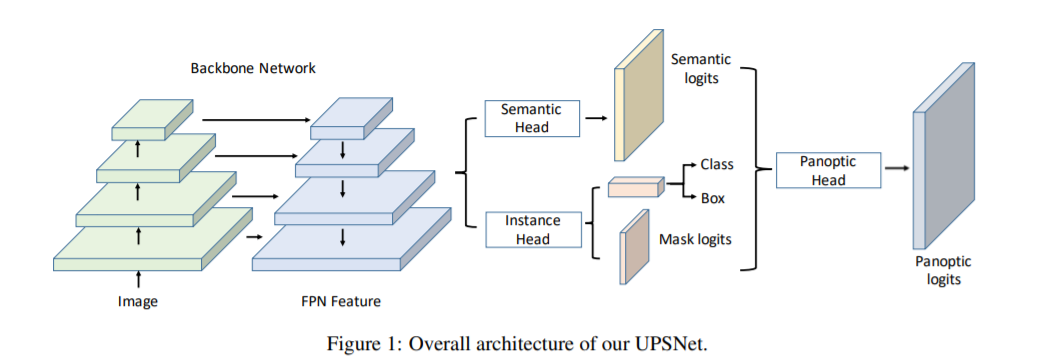
Xiong, Yuwen, et al. "UPSNET: A unified panoptic segmentation network." Proceedings of the IEEE

Conference on Computer Vision and Pattern Recognition. 2019.

1. Citation = 60
2. CityScapes Ranking:
   * 1. Panoptic Semantic Labeling Task [None]
     2. Instance-Level Semantic Labeling Task No. 15

First presented: Conference on Computer Vision and Pattern Recognition. 2019.

1. Top-down / Proposal-based approach: the authors designed two heads on top of the backbone for solving the semantic labelling tasks for semantic and instance segmentations simultaneously
2. Briefly describe and sketch the full architecture of the proposed model.



* The UPSNet consists of a shared convolutional **feature extraction backbone** and multiple heads on top of it. Each head is a sub-network which leverages the features from the backbone and serves a specific design purpose
* The original **Mask R-CNN backbone** is adopted as UPSNet’s convolutional feature extraction network.
* Reference: K. He, G. Gkioxari, P. Dollar, and R. Girshick. Mask r-cnn. In ICCV, 2017

1. If the paper is about a top-down approach, briefly describe how object proposals are generated: Does the model predict anchor boxes? Is one instance mask predicted for each proposal? Is the object proposal method novel or was it introduced in a previous paper? If so, give a reference to this paper

* The panoptic head is designed to predict the final panoptic segmentation via pixel-wise classification. It exploits the logits from the two heads (i.e. the semantic head and the instance head) and adds a new channel of logits corresponding to an extra unknown class. By doing so, it provides a better way of resolving the conflicts between semantic and instance segmentation.
* Yes, bounding boxes are predicted
* Object proposal method follows the Mask R-CNN design (see reference above)

1. Not applicable.

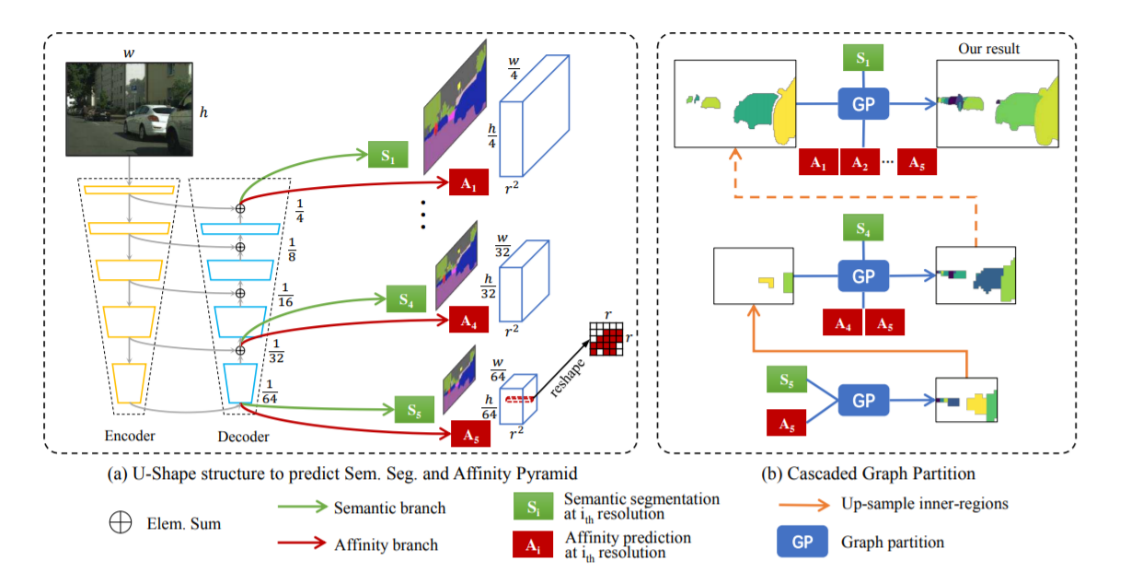
**Computer Vision Foundations – Exercise 5 – Zhao Sun**

Gao, Naiyu, et al. "SSAP: Single-shot instance segmentation with affinity pyramid." Proceedings of the IEEE International Conference on Computer Vision. 2019.

1. Citation = 19
2. CityScapes Ranking:
   * 1. Panoptic Semantic Labeling Task No. 11
     2. Instance-Level Semantic Labeling Task No. 17

First presented: International Conference on Computer Vision. 2019.

1. Bottom-up / Proposal-free approach: this work proposes a single-shot proposal-free instance segmentation method that requires only one single pass for prediction, in which an affinity pyramid (i.e., the pixel-pair affinity computing the probability that two pixels belong to the same instance) can be jointly learn with the labelling of semantic class.
2. Briefly describe and sketch the full architecture of the proposed model.



* The model consists of two parts: (a) a unified U-shape network to learn the semantic segmentation and affinity pyramid with a single backbone network, and (b) a cascaded graph partition module to sequentially generate multi-scale instance predictions using the jointly learned affinity pyramid and semantic segmentation.

1. If the paper is about a bottom-up approach, briefly describe how the final instance segmentation is obtained from the prediction of the model: Is the approach using some kind of post-processing or clustering algorithm? If so, which one?

* Different object instances are distinguished based on the **instance-aware pixel-pair affinity**, which specifies whether two pixels belong to the same instance or not; both short-range and long-range affinities are used in the learning resulting in the multi-scale affinity pyramid
* Graph partition based post-processing algorithm is used here. More specifically, a graph is constructed by regarding each pixel as a node and transforming affinities into the edge scores. Graph partition is then employed from higher-level lower-resolution layers to lower-level higher-resolution layers progressively.