Combinatorial Optimization for Panoptic Segmentation: A Fully Differentiable Approach

Ahmed Abbas, Paul Swoboda

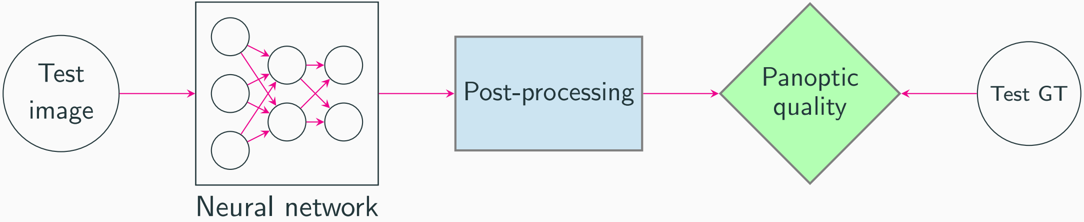
Max Planck Institute for Informatics, Germany

Panoptic segmentation = Semantic / Instance seg.

$\longrightarrow$

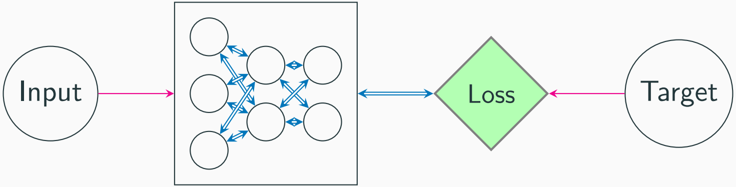
Kirillov 2019

Evaluation pipeline for panoptic segmentation



Conventional training pipeline

- Post-processing block not part of training
- Test-time evaluation metric not used in training



Motivation

Existing methods

1. Do not optimize for panoptic labels
2. Require many parameters (e.g. loss balancing weights, post-processing)
3. Complex architecture (e.g. additional Mask-RCNN for ROI)

Motivation

Existing methods

1. Do not optimize for panoptic labels
2. Require many parameters (e.g. loss balancing weights, post-processing)
3. Complex architecture (e.g. additional Mask-RCNN for ROI)

Our aims

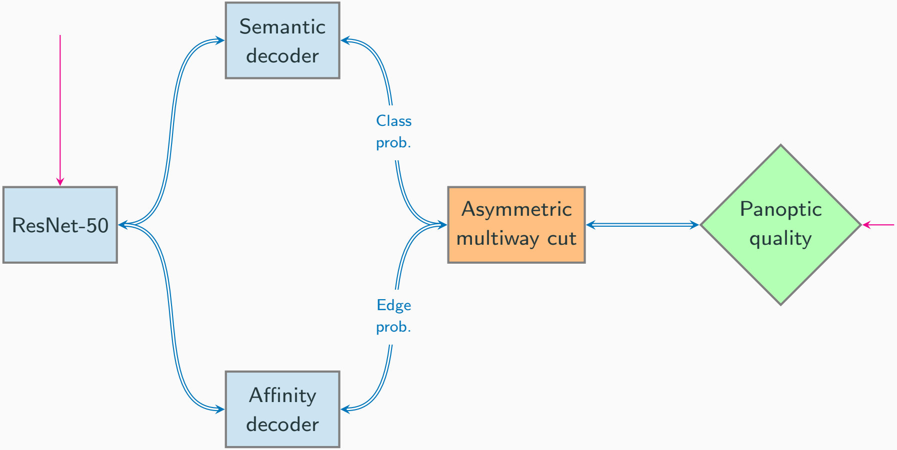
1. Fully differentiable pipeline
2. Optimize for the evaluation metric
3. Fewer hyperparameters
4. Use ‘standard’ neural network architecture (ResNet50)

Literature review

Methods	Simplicity	Params.		End-to-end	Optimize PQ	Performance
		Train	Eval			
MaxDeepLab[1]						
E cientPS[2]						
AxialDeepLab[3]						
PanopticDeepLab[4]						
UPSNet[5]						
SMW[6]						
SSAP[7]						
Our aim						?

[1]Wang 2020b, [2]Mohan 2021, [3]Wang 2020a, [4]Cheng 2020, [5]Xiong 2019, [6]Wolf 2020, [7]Gao 2019

Our pipeline



Panoptic labels

Multiway cut (MWC) - Calinescu 2008

Generalization of graph-cut on $G = (V, E)$ for $K > 2$

$$
\begin{aligned}
\min_{\substack{x: V \to \{1, \dots, K\}, \\ y: E \to \mathbb{B} \cap \{0,1\}}} \quad & \sum_{i \in V} c_V(i, x(i)) + \sum_{ij \in E} c_E(ij) y(ij) \\
\text{s.t.} \quad & y(ij) = 0, \ \text{if } x(i) = x(j) \\
& y(ij) = 1, \ \text{if } x(i) \neq x(j)
\end{aligned}
$$

Multiway cut (MWC)

Generalization of graph-cut on $G = (V, E)$ for $K > 2$

$$\min_{\substack{x: V \to \{1,\dots,K\}, \\ y: E \to B \cap \{0,1\}}} \sum_{i \in V} c_V(i, x(i)) + \sum_{ij \in E} c_E(ij) y(ij)$$

$$\text{s.t.} \quad \begin{aligned} y(ij) &= 0, \text{ if } x(i) = x(j) \\ y(ij) &= 1, \text{ if } x(i) \neq x(j) \end{aligned}$$

c_V, c_E : Semantic, edge costs

$x(i)$: Semantic label of i in V

$$y(ij) = \begin{cases} 0, & i, j \text{ belong to same instance} \\ 1, & i, j \text{ belong to different instance} \end{cases}$$

B : Enforce valid clustering

Multiway cut (MWC)

Generalization of graph-cut on $G = (V, E)$ for $K > 2$

$$\min_{\substack{x: V \to \{1, \dots, K\}, \\ y: E \to B \cap \{0,1\}}} \sum_{i \in V} c_V(i, x(i)) + \sum_{ij \in E} c_E(ij) y(ij)$$

$$\text{s.t.} \quad y(ij) = 0, \text{ if } x(i) = x(j)$$
$$y(ij) = 1, \text{ if } x(i) \neq x(j)$$

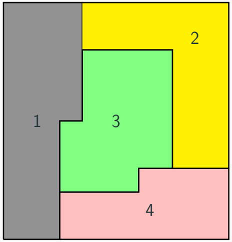
c_V, c_E : Semantic, edge costs

$x(i)$: Semantic label of i in V

$$y(ij) = \begin{cases} 0, & i, j \text{ belong to same instance} \\ 1, & i, j \text{ belong to different instance} \end{cases}$$

B : Enforce valid clustering

$K = 4$



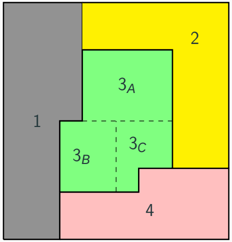
Asymmetric multiway cut (AMWC) - Kroeger 2014

AMWC on graph $G = (V, E)$

$$\min_{\substack{x:V \to \{1,\dots,K\},\\ y:E \to B \cap \{0,1\}}} \sum_{i \in V} c_V(i, x(i)) + \sum_{ij \in E} c_E(ij) y(ij)$$

$$\text{s.t.} \quad y(ij) = 0, \ \text{if } x(i) = x(j) \notin P$$
$$y(ij) = 1, \ \text{if } x(i) \neq x(j)$$

$K = 4, P = \{3\}$



c_V, c_E : Semantic, edge costs

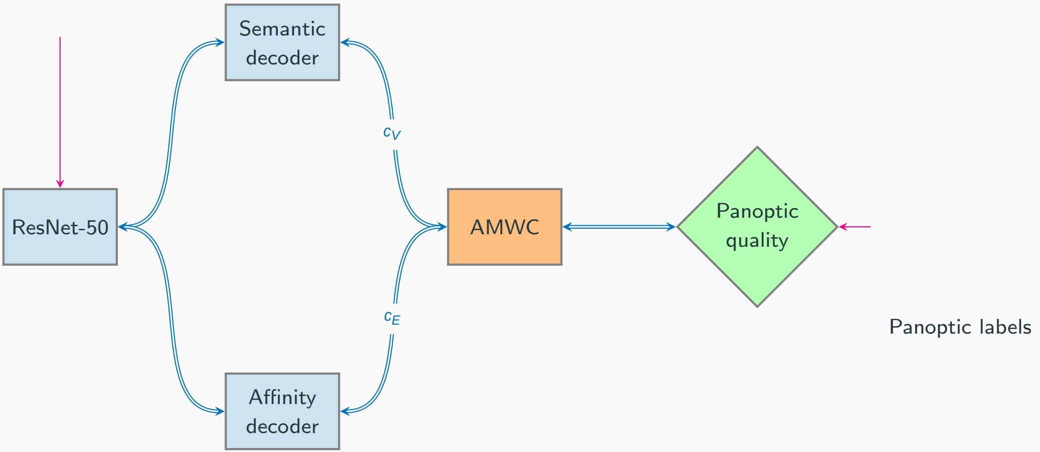
B : Enforce valid clustering

$P \subseteq [K]$: Partitionable classes

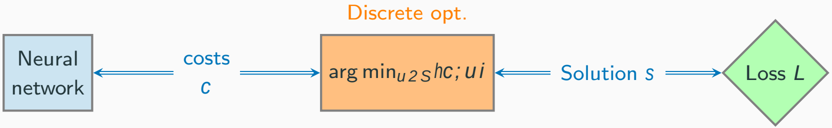
$x(i)$: Semantic label of i in V

$$y(ij) = \begin{cases} 0, & i, j \text{ belong to same instance} \\ 1, & i, j \text{ belong to different instance} \end{cases}$$

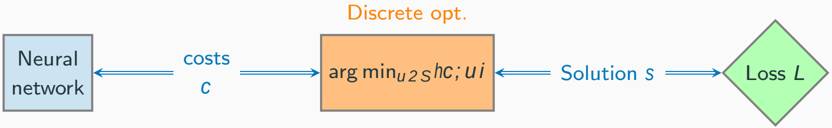
Recap: our pipeline



Gradient estimation through discrete optimization layers



Gradient estimation through discrete optimization layers



Build on the approach of Vlastelica 2019

$$\frac{\partial L}{\partial c} = \frac{1}{\ } \quad s(c + \quad \frac{\partial L}{\partial s}) \quad s(c)$$

Gradient estimation through discrete optimization layers

Build on the approach of Vlastelica 2019

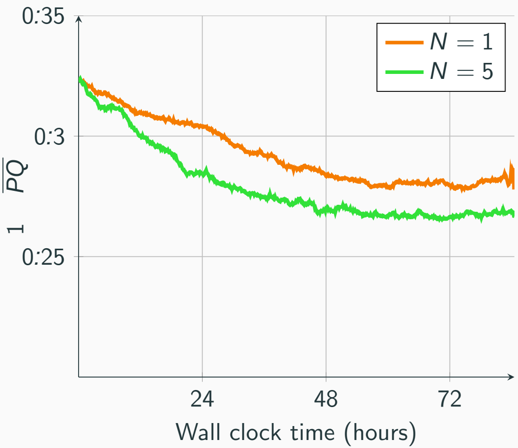
$$\frac{\partial L}{\partial c} = \frac{1}{\lambda}\left[s\left(c + \lambda \frac{\partial L}{\partial s}\right) - s(c)\right]$$

Our extension

$$\frac{\partial L_{avg}}{\partial c} = \frac{1}{N} \sum_{i}^{N} \frac{\partial L^{\lambda_i}}{\partial c}, \qquad \lambda_i \sim U(a, b)$$

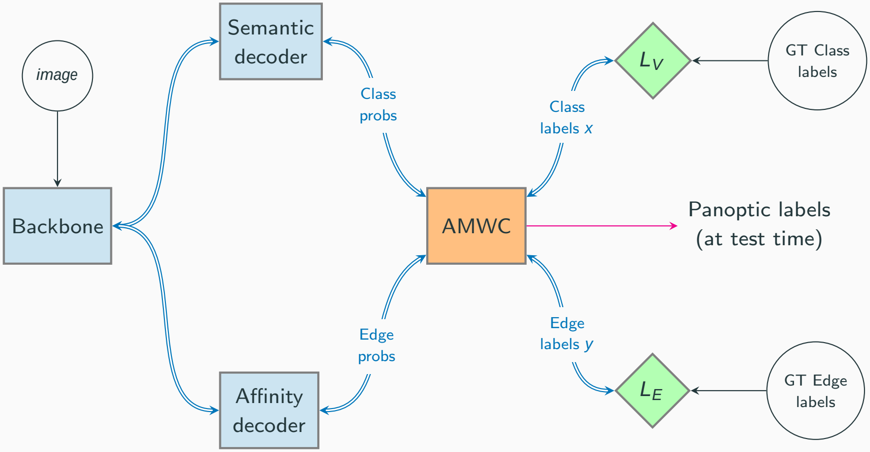
- Averages gradients over multiple scales
- 3 faster convergence

Contribution 1: Improved gradient estimation



Training loss comparison ($N = 5$: our extension with **improved convergence**)

Panoptic Segmentation: Naïve fully differentiable pipeline



Naïve pipeline: Gradient estimation through AMWC

Perturb semantic, edge costs by respective gradients

$$
\begin{aligned}
\min_{\substack{x:V \to \{1,\dots,K\},\\ y:E \to B \cap \{0,1\}}} \quad & \sum_i \left[c_V(i,x(i)) + \frac{\partial L_V}{\partial x}((i,x(i)) \right] + \sum_{ij} \left[c_E(ij) + \frac{\partial L_E}{\partial y}(ij) \right] y(ij) \\
\text{s.t.} \quad & y(ij) = 0, \ \text{if } x(i) = x(j) \notin P \\
& y(ij) = 1, \ \text{if } x(i) \neq x(j)
\end{aligned}
$$

Panoptic Segmentation: Naïve fully differentiable pipeline

Does not perform well:

- Edge misclassifications are **not** equally important
- At test-time we care about pixel labels not edge labels
- Need to optimize test-time metric of panoptic quality (PQ)

Edge labels: ground-truth, prediction, correct predictions

Contribution 2: Optimize panoptic quality surrogate

Approximate non-differentiable panoptic quality metric

$$PQ = \frac{\sum_{(p,g)\in TP} IoU(p,g)}{|TP| + 0.5(|FP| + |FN|)}$$

by

$$\overline{PQ} = \frac{\sum_{(p,g)\in\overline{TP}} h(IoU(p,g)) \quad (p)\, IoU(p,g)}{\sum_{(p,g)\in\overline{TP}} h(IoU(p,g)) \quad (p) + 0.5\{\sum_{p\in\overline{FP}} \quad (p) + |\overline{FN}|\}}$$

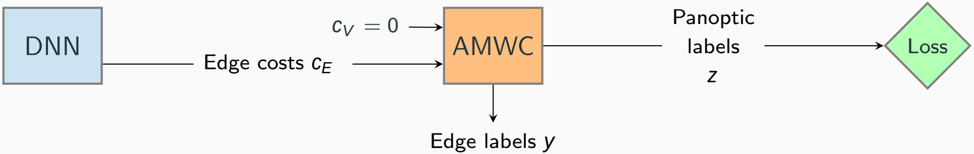
where

$h(\ )$: Soft-thresholding

$\ (p)$: Foreground probability (from mask area)

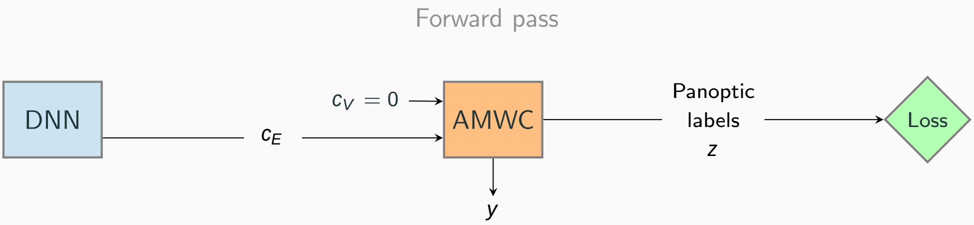
Contribution 3: Backprop with loss on panoptic labels ($K = 1$)

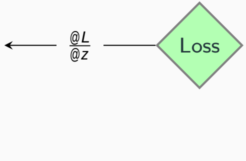
Forward pass



Ignored semantic costs c_V for brevity

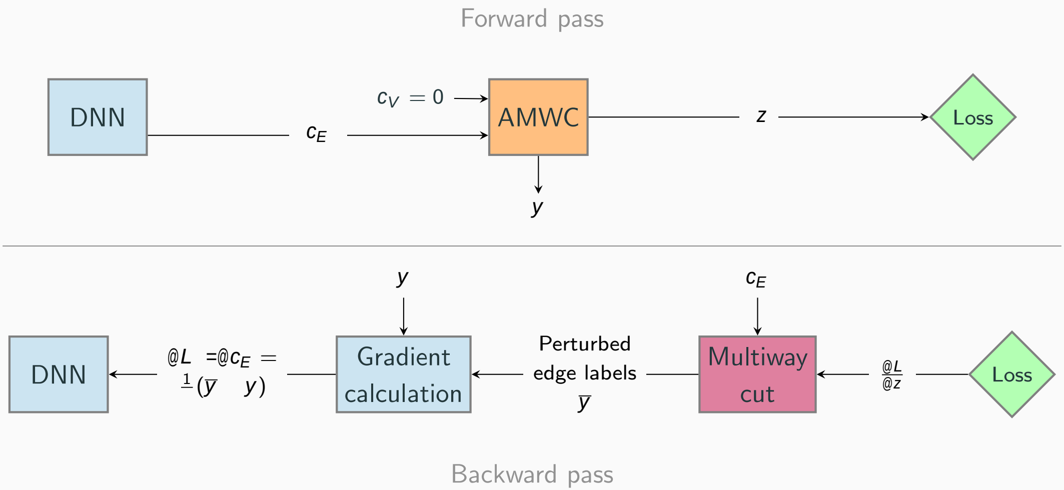
Contribution 3: Backprop with loss on panoptic labels ($K = 1$)





Ignored semantic costs c_V for brevity

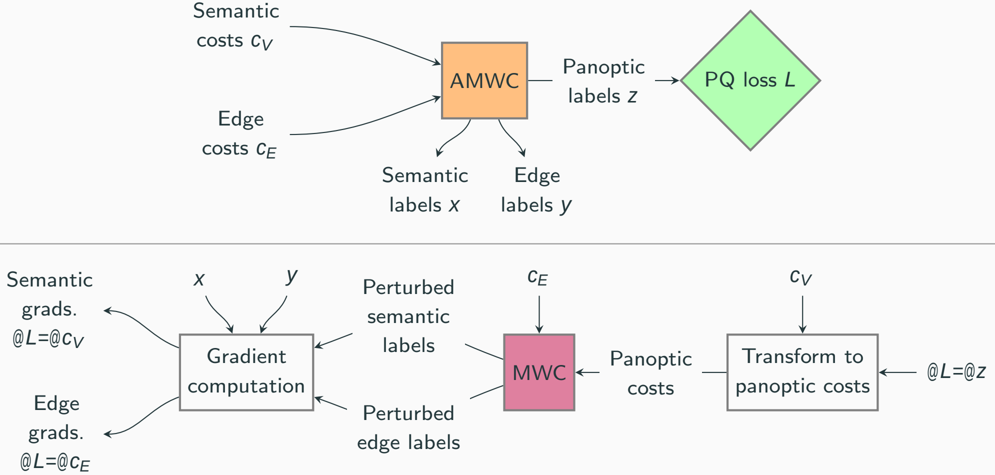
Contribution 3: Backprop with loss on panoptic labels ($K = 1$)



Ignored semantic costs c_V for brevity

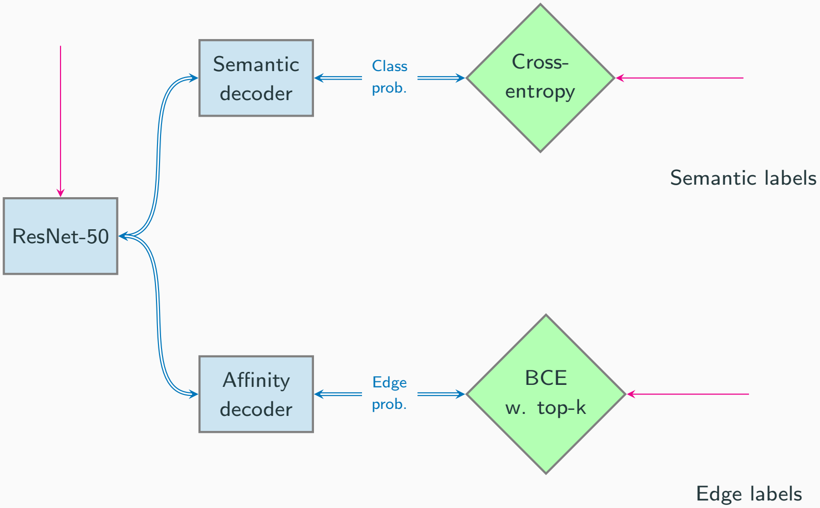
Contribution 3: Full backprop (K 1) with loss on panoptic labels

Forward pass



Backward pass with transformation to panoptic label space

Baseline: train by 'usual' losses



Results

- Outperform all comparable approaches
- Sometimes, even with a disadvantage (e.g. smaller backbone)

Methods	Architecture	Hyperparameters		E-to-E	Opt. PQ	Panoptic qual.	
		Train	Eval			Citysc.	COCO
MaxDeepLab						-	49.3
E cientPS						63.9	-
AxialDeepLab						63.9	41.8
UPSNet						59.3	42.5
SSAP						61.1	36.5
SMW						59.3	-
PanopticDeepLab						60.2	35.1
Our baseline						58.5	34.3
Our fully differentiable						**62.1**	**38.4**

Summary

- Simple and fully differentiable pipeline for panoptic segmentation
- Improved gradient estimation through discrete optimization problems
- Optimize panoptic quality differentiable surrogate
- Transformation to MWC in backward pass to compute gradients
- First large-scale study of backpropagation through heuristic optimization solvers
- *Shortcoming:* Inference time of around $2s$ per image
- Code available at: `github.com/aabbas90/COPS`

References i

- Calinescu, Gruia (2008). "Multiway Cut". In: *Encyclopedia of Algorithms*. Ed. by Ming-Yang Kao. Boston, MA: Springer US, pp. 567–569. ISBN: 978-0-387-30162-4. DOI: 10.1007/978-0-387-30162-4_253. URL: https://doi.org/10.1007/978-0-387-30162-4_253.
- Cheng, Bowen (2020). "Panoptic-deeplab: A simple, strong, and fast baseline for bottom-up panoptic segmentation". In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 12475–12485.
- Gao, Naiyu (2019). "Ssap: Single-shot instance segmentation with affinity pyramid". In: *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pp. 642–651.
- Kirillov, Alexander (2019). "Panoptic segmentation". In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 9404–9413.

Kroeger, Thorben (2014). "Asymmetric cuts: Joint image labeling and partitioning". In: *German Conference on Pattern Recognition*. Springer, pp. 199–211.

Mohan, Rohit (2021). "Efficientps: Efficient panoptic segmentation". In: *International Journal of Computer Vision*, pp. 1–29.

Vlastelica, Marin (2019). "Differentiation of blackbox combinatorial solvers". In: *International Conference on Learning Representations*.

Wang, Huiyu (2020a). "Axial-deeplab: Stand-alone axial-attention for panoptic segmentation". In: *European Conference on Computer Vision*. Springer, pp. 108–126.

Wang, Huiyu (2020b). "MaX-DeepLab: End-to-End Panoptic Segmentation with Mask Transformers". In: *arXiv preprint arXiv:2012.00759*.

- Wolf, Steffen (2020). "The Semantic Mutex Watershed for Efficient Bottom-Up Semantic Instance Segmentation". In: *European Conference on Computer Vision*. Springer, pp. 208–224.
- Xiong, Yuwen (2019). "Upsnet: A unified panoptic segmentation network". In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 8818–8826.