Computer Vision I

Jannik Irmai, Jannik Presberger, Bjoern Andres

Machine Learning for Computer Vision
TU Dresden

https://mlcv.cs.tu-dresden.de/courses/24-winter/cv1/

Winter Term 2024/2025

So far, we have studied

- **pixel classification**, a problem whose feasible solutions define decisions at the pixels of an image
- **image decomposition**, a problem whose feasible solutions decide whether pairs of pixels are assigned to the same or distinct components of the image.

Applications exists for which both problems are too restrictive:

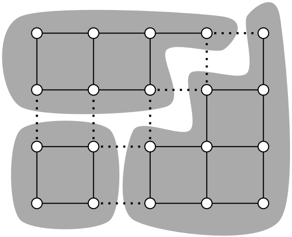
- In **pixel classification**, there is no way of assigning neighboring pixels of the same class to distinct components of the image.
- In **image decomposition**, there is no way of expressing that a unique decision shall be made for pixels that belong to the same component of the image.

Semantic segmentation

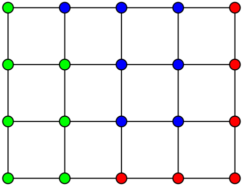
M. Cordts, M. Omran, S. Ramos, T. Rehfeld, M. Enzweiler, R. Benenson, U. Franke, S. Roth, and B. Schiele. The Cityscapes Dataset for Semantic Urban Scene Understanding. CVPR 2016. See also: https://www.cityscapes-dataset.com/

- One application where a joint generalization of pixel classification and image decomposition is useful is called **semantic segmentation**.
- In the above image, thin boundaries are left between pixels of the same class (e.g. pedestrian) that belong to different instances of the class (e.g. distinct pedestrians).
- Next, we are going to introduce a strict generalization of both, pixel classification and image decomposition that does not require these boundaries.

Semantic segmentation



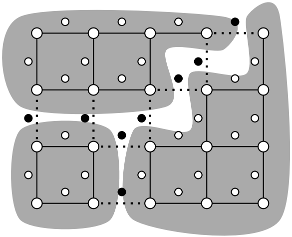
Graph Decomposition



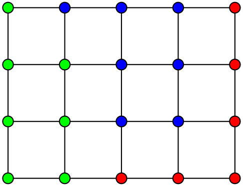
Node Labeling

We state an optimization problem whose feasible solutions define both, a **decomposition** of a graph $G = (V, E)$ and a **labeling** $l\colon V \to L$ of its nodes.

Semantic segmentation



Graph Decomposition

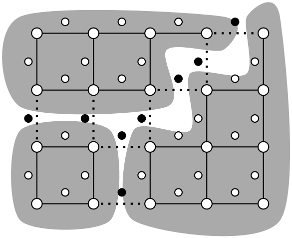


Node Labeling

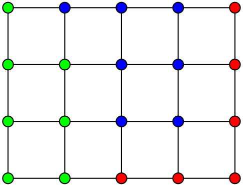
We encode feasible decompositions by multicuts:

$$Y_G := \left\{ y : E \to \{0,1\} \;\middle|\; \forall C \in \mathsf{cycles}(G)\, \forall e \in C\colon\; y_e \le \sum_{f \in C \setminus \{e\}} y_f \right\}$$

Semantic segmentation

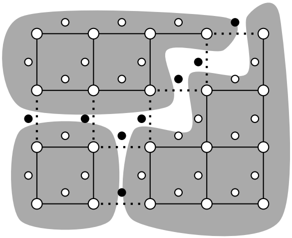


Graph Decomposition

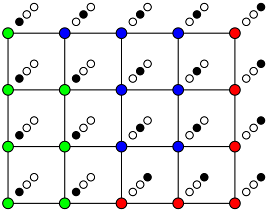


Node Labeling

Semantic segmentation



Graph Decomposition

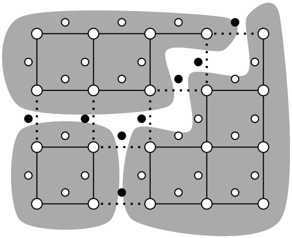


Node Labeling

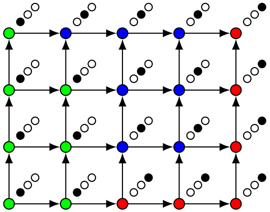
We encode feasible node labelings by binary vectors:

$$Z_{VL} := \left\{ z : V \times L \to \{0,1\} \;\middle|\; \forall v \in V : \sum_{l \in L} z_{vl} = 1 \right\}$$

Semantic segmentation



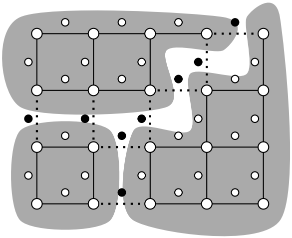
Graph Decomposition



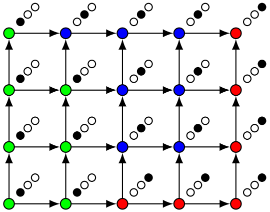
Node Labeling

We choose an arbitrary orientation (V, A) of the edges E, i.e., for each $v, w \in V$, we have $\{v, w\} \in E$ if and only if either $(v, w) \in A$ or $(w, v) \in A$.

Semantic segmentation



Graph Decomposition



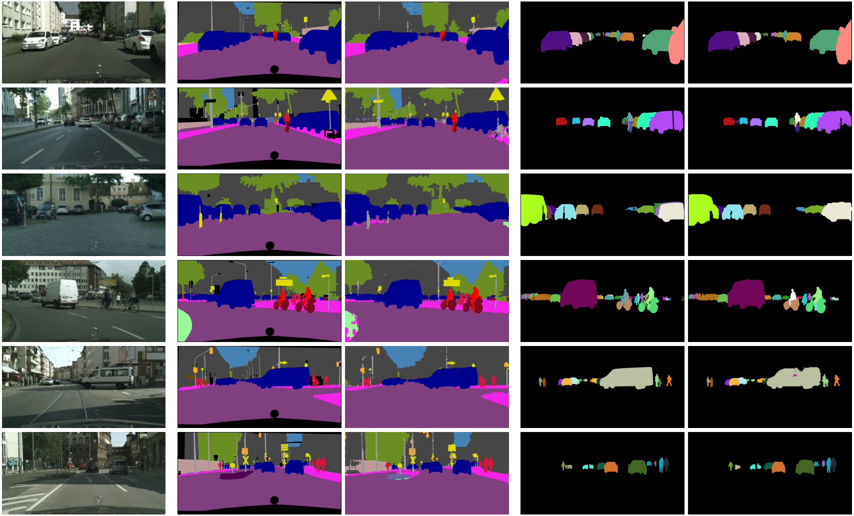
Node Labeling

W.r.t. the orientation (V, A) of the graph $G = (V, E)$, the set L of labels, any (costs) $c\colon V \times L \to \mathbb{R}$ and any (costs) $c', c''\colon A \times L^2 \to \mathbb{R}$, the instance of the **joint graph decomposition and node labeling problem** has the form

$$\min_{(y,z) \in Y_G \times Z_{VL}} \sum_{v \in V} \sum_{l \in L} c_{vl}\, z_{vl} + \sum_{(v,w) \in A} \sum_{(l,l') \in L^2} c'_{vwll'}\, z_{vl}\, z_{wl'}\, y_{\{v,w\}} + \sum_{(v,w) \in A} \sum_{(l,l') \in L^2} c''_{vwll'}\, z_{vl}\, z_{wl'}\, (1 - y_{\{v,w\}})$$

Semantic segmentation

Image	Labeling		Decomposition	
	Truth	Prediction	Truth	Prediction



Kirillov et al. 2017. doi:10.1109/CVPR.2017.774