

Computer Vision I

Jannik Presberger, David Stein, Bjoern Andres

Machine Learning for Computer Vision
TU Dresden



<https://mlcv.cs.tu-dresden.de/courses/25-winter/cv1/>

Winter Term 2025/2026

Object recognition

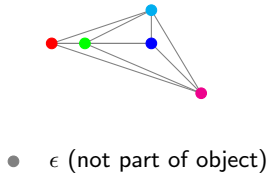
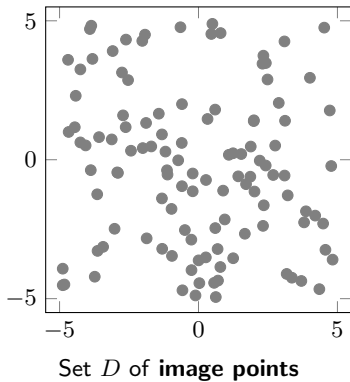
Object recognition is the task of finding any occurrences of an object in an image, given a **model** of the the geometry and appearance of the object.

Object recognition



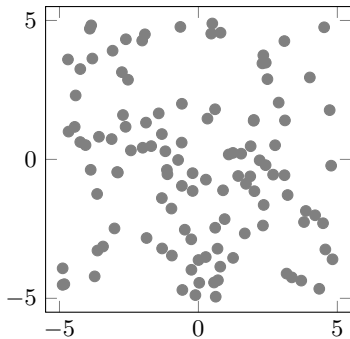
Pishchulin, Insafutdinov, Tang, A, Andriluka, Gehler, Schiele 2016.
Insafutdinov, Pishchulin, A, Andriluka, Schiele 2016.

Object recognition

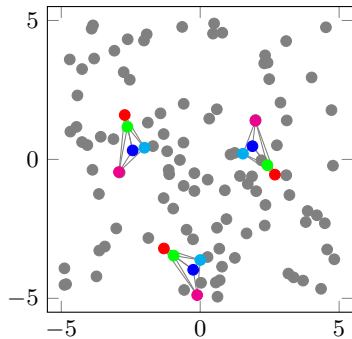


Set V of object points

Object recognition



Set D of **image points**



Recognition

Decisions at points

- ▶ For any **image point** $d \in D$ and any **object point** $v \in V$, let $z_{dv} \in \{0, 1\}$ indicate whether d is an occurrence of v .
- ▶ We constrain each image point to be an occurrence of precisely one object point, possibly ϵ . Hence, we consider the feasible set

$$Z_{DV} = \left\{ z: D \times V \rightarrow \{0, 1\} \mid \forall d \in D: \sum_{v \in V} z_{dv} = 1 \right\} .$$

Costs at points

- ▶ For any image point $d \in D$ and any object point $v \in V$, let $c_{dv} \in \mathbb{R}$ a cost associated with the decision $y_{dv} = 1$
- ▶ This cost typically depends on the contents of the image at the point d .
- ▶ It can be estimated from examples by the machine learning techniques discussed earlier.

Decisions for pairs of points

- For any pair $\{d, d'\} \in \binom{D}{2}$ of image points, let $y_{\{d, d'\}} \in \{0, 1\}$ indicate whether d and d' belong to the same occurrence of an object in the image
- We require these decisions to be transitive, i.e.

$$\forall d \in D \ \forall d' \in D \setminus \{d\} \ \forall d'' \in D \setminus \{d, d'\}: \\ y_{\{d, d'\}} + y_{\{d', d''\}} - 1 \leq y_{\{d, d''\}} \quad (1)$$

Hence, we consider the feasible set

$$Y_D = \{y: \binom{D}{2} \rightarrow \{0, 1\} \mid (1)\}$$

Costs for pairs of points

- ▶ For any pair $(d, d') \in D^2$ of image points such that $d \neq d'$ and any pair $(v, w) \in V^2$ of object points, let
 - ▶ $c'_{dd'vw} \in \mathbb{R}$ a cost associated with the decision $z_{dv} z_{d'w} y_{\{d,d'\}} = 1$
 - ▶ $c''_{dd'vw} \in \mathbb{R}$ a cost associated with the decision $z_{dv} z_{d'w} (1 - y_{\{d,d'\}}) = 1$
- ▶ These costs can depend, e.g., on the distance between d and d' in the image plane.
- ▶ They can be estimated from examples by the machine learning techniques discussed earlier.

Optimization problem

- The task of object recognition can now be stated as the optimization problem

$$\begin{aligned} \min_{(y,z) \in Y_D \times Z_{DV}} & \sum_{d \in D} \sum_{v \in V} c_{dv} z_{dv} \\ & + \sum_{d \in D} \sum_{d' \in D \setminus \{d\}} \sum_{(v,w) \in V^2} c'_{dd'vw} z_{dv} z_{d'w} y_{\{d,d'\}} \\ & + \sum_{d \in D} \sum_{d' \in D \setminus \{d\}} \sum_{(v,w) \in V^2} c''_{dd'vw} z_{dv} z_{d'w} (1 - y_{\{d,d'\}}) \end{aligned}$$

- This is a joint graph decomposition and node labeling problem.
- The same local search algorithms we have considered for the task of semantic segmentation can be applied.