Neighbourhood Consensus Networks

**Introduction**

**Task:** find pixel-level visual correspondences

**Challenges:** strong illumination or appearance changes, changes across time, intra-class variation, repetitive structures

**Contributions:**
- Trainable method for feature extraction, matching and filtering
- Based on dense CNN descriptors and a novel 4D neighbourhood consensus CNN

**Review - classical pipeline:**

**Proposed method**

**Feature extraction and matching**

**Trainable match filtering**

**Neighbourhood Consensus Network**

4-D conv. layers

The network is applied twice for invariance to pair order:

\[
\text{filters of the first layer span is a 4D CNN}
\]

Training loss:

The network is trained with weak supervision:

\[
\mathcal{L}^{+}(p) = \mathcal{L}^{-}(p) = \lVert p + \lambda \rVert^2
\]

**Extracting correspondences:**

Matches can be extracted in both directions from the output \( \mathbf{J} \):

**Experimental results**

**Category-level matching:** PF-Pascal dataset [1]

- **Task:** match similar semantic parts
- **Metric:** percentage of correctly keypoints (PCK)

**Instance-level matching:** InLoc dataset [5]

- **Task:** retrieve 6-dof camera poses of query images
- **Metric:** percentage of correctly localized queries

**References**


