Convolutional neural network architecture for geometric matching

Ignacio Rocco[1,2] **Relja Arandjelović**[1,2,*] **Josef Sivic**[1,2,3]

[1]DI ENS, École normale supérieure, PSL Research University [2]INRIA [3]CIIRC, CTU in Prague

*Now at DeepMind

Goal

- Instance-level and category-level image alignment
  - **Output:** smooth dense correspondence field

Challenges

- Substantial appearance differences
- Presence of background clutter
- Lack of large annotated real image pair dataset

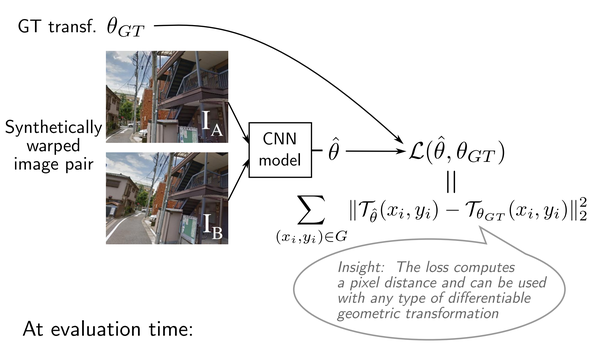
Contributions

- CNN architecture suitable for category-level image alignment
- The model is trainable from synthetically warped image pairs
- Matching layer enables generalization to real image pairs

Overview

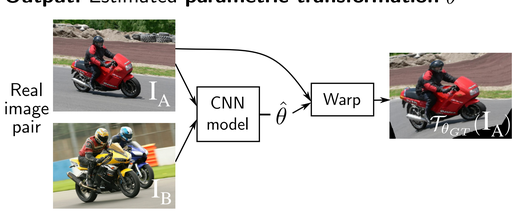
At training time:

- **Inputs:** - Synthetically warped image pair
  - **ground truth transf.** θ_{GT}
- **Output:** Estimated **parametric transformation** $\hat{\theta}$



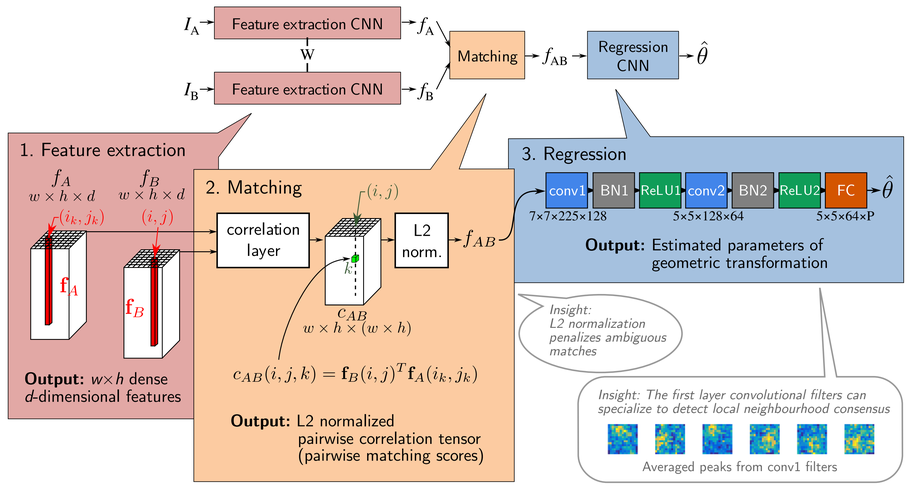
At evaluation time:

- **Input:** Real image pair
- **Output:** Estimated **parametric transformation** $\hat{\theta}$



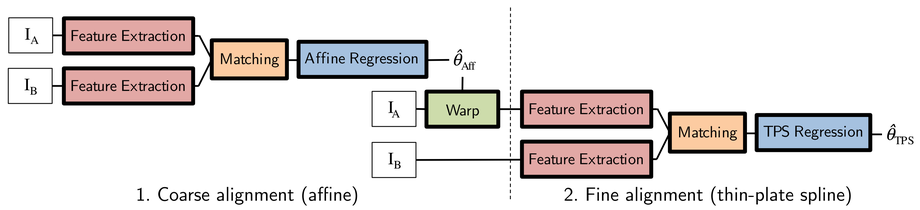
Model

- Three stage siamese CNN architecture mimicking the classical matching pipeline
  1. **Feature extraction CNN:** pre-trained VGG-16 model + per-column L2-normalization
  2. **Matching:** correlation layer + per-column L2-normalization
  3. **Regression CNN:** small CNN, trained from scratch



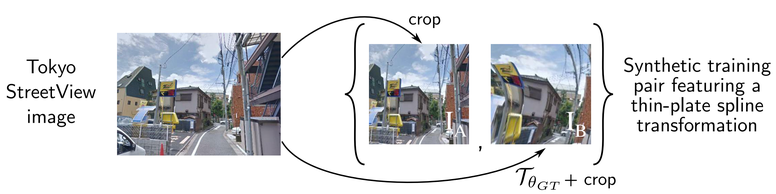
Coarse-to-fine matching architecture

- The same architecture can be applied with increasing geometric model complexity
  1. Coarse alignment using an **affine transformation**
  2. Refined alignment using a **thin-plate spline transformation**
- The final transformation is the composition of both stages



Training from synthetic imagery

- **Training pairs:** generated by a real and a synthetically warped image



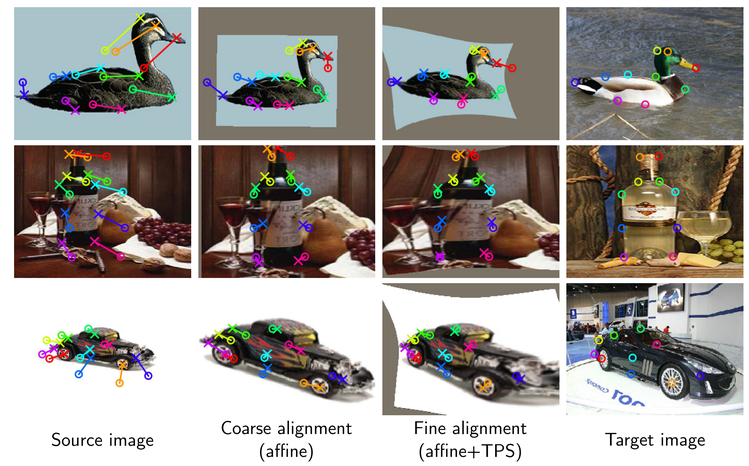
- **Generalization:** We show that the method is relatively unaffected by the nature of the training images

Results on the Proposal Flow dataset

- Evaluated using annotated keypoints
- Metric: Percentage of correct keypoints (PCK)

Methods	PCK (%)
DeepFlow [1]	20
GMK [2]	27
SIFT Flow [3]	38
DSP [4]	29
Proposal Flow [5]	56
RANSAC with our features (affine)	47
Ours (affine)	49
Ours (affine + thin plate spline)	56
Ours (affine ensemble + thin plate spline)	**57**

Qualitative results:

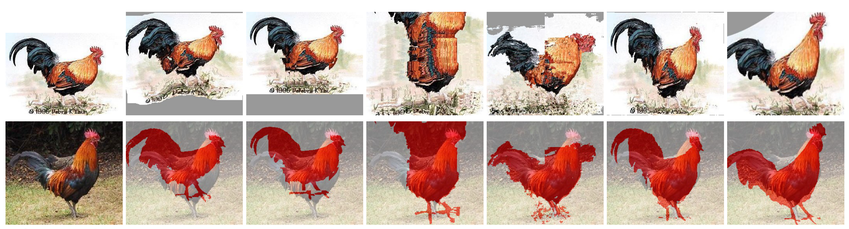


Results on the Caltech-101 dataset

- Evaluated using annotated object segmentation masks
- Metrics: Label transfer accuracy (LT-ACC), Intersection over union (IoU), Localization error (LOC-ERR)

Methods	LT-ACC	IoU	LOC-ERR
DeepFlow [1]	0.74	0.40	0.34
GMK [2]	0.77	0.42	0.34
SIFT Flow [3]	0.75	0.48	0.32
DSP [4]	0.77	0.47	0.35
Proposal Flow [5]	0.78	0.50	**0.25**
Ours (affine)	0.79	0.51	**0.25**
Ours (affine + thin-plate spline)	**0.82**	**0.56**	**0.25**

Qualitative comparison to other methods:



References

[1] P. Weinzaepfel, et al. DeepFlow: Large displacement optical flow with deep matching. In Proc. ICCV, 2013

[2] O. Duchenne, et al. A graph-matching kernel for object categorization. In Proc. ICCV, 2011

[3] C. Liu, et al. SIFT Flow: Dense correspondence across scenes and its applications. IEEE PAMI, 2011

[4] J. Kim, et al. Deformable spatial pyramid matching for fast dense correspondences. In Proc. CVPR, 2013

[5] B. Ham, M. Cho, C. Schmid, and J. Ponce. Proposal Flow. In Proc. CVPR, 2016