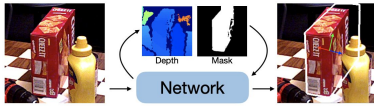
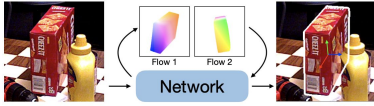


Problems

- 6D object annotations are hard to obtain or prone to contain large labeling errors.
- Existing self-supervised methods rely on additional supervision signal, such as depth or masks, which are cumbersome to obtain.

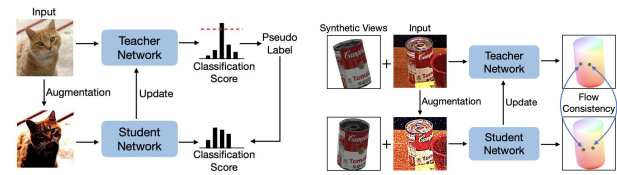


(a) Existing self-supervised 6D object pose methods



(b) Our solution based on pseudo flow consistency

- The two fundamental problems of pseudo labeling are critical yet open questions for 6D object pose estimation:
 - the generative strategy of creating pseudo labels
 - the selection strategy of extracting high-quality labels



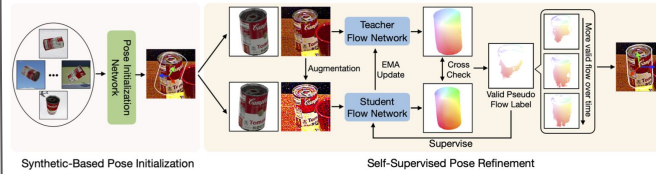
(a) The standard strategy for self-supervised classification (b) The proposed strategy for self-supervised object pose estimation

Code is available at:

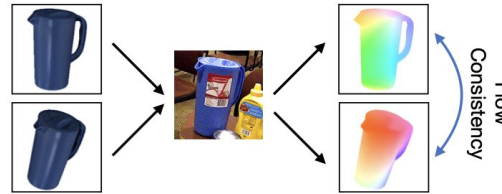
<https://github.com/YangHai-1218/PseudoFlow>

Solution

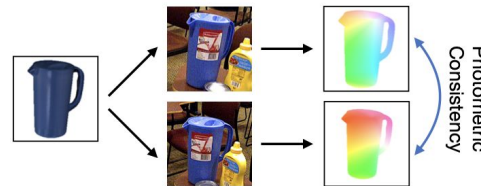
- Overall framework
 - Train a pose estimation model on synthetic images.
 - Refinement framework built upon optical flow estimation is trained on un-annotated real images.



- Flow Consistency across multiple views for high-quality pseudo label selection.



- Flow-guided photometric consistency further improves the robustness.



Experiments

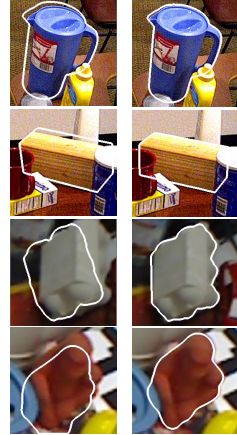
Ablation study:

\mathcal{L}_{flow}	\mathcal{L}_{photo}	MSPD	MSSD	VSD	ADD
-	-	0.759	0.589	0.519	30.9
-	✓	0.765	0.631	0.578	37.5
✓	-	0.780	0.711	0.658	64.2
✓	✓	0.785	0.749	0.664	67.4

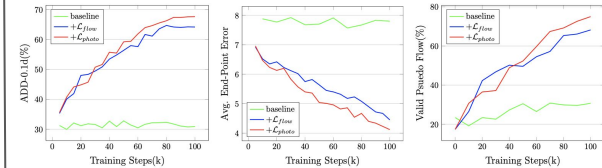
Close to performance trained with real annotations:

Method	MSPD	MSSD	VSD	ADD
Initialization v1	0.632	0.491	0.420	27.4
+ Ours (Real)	0.780	0.731	0.673	64.6
+ Ours (SSL)	0.759	0.722	0.650	63.2
Initialization v2	0.673	0.580	0.508	36.0
+ Ours (Real)	0.775	0.722	0.660	65.3
+ Ours (SSL)	0.764	0.724	0.643	64.2
Initialization v3	0.694	0.598	0.522	38.6
+ Ours (Real)	0.803	0.752	0.686	69.2
+ Ours (SSL)	0.785	0.749	0.664	67.4

Qualitative results:



Training analysis:



Initialization +Ours(SSL)

Visualization with increasing training iterations:

