Viewpoint Estimation for Workpieces with Deep Transfer Learning from Cold To Hot

Changsheng Lu, Haoitan Wang, Chaochen Gu, Kaijie Wu, Xinping Guan

Presenter: Changsheng Lu
Shanghai Jiao Tong University
Shanghai, China

Dec. 13, Siem Reap, Cambodia
ICONIP 2018

Outline

- Introduction
- Deep transfer networks with cold-to-hot training strategy
- Experimental results
- Conclusion

Outline

- **Introduction**
- **Deep transfer networks with cold-to-hot training strategy**
- **Experimental results**
- **Conclusion**

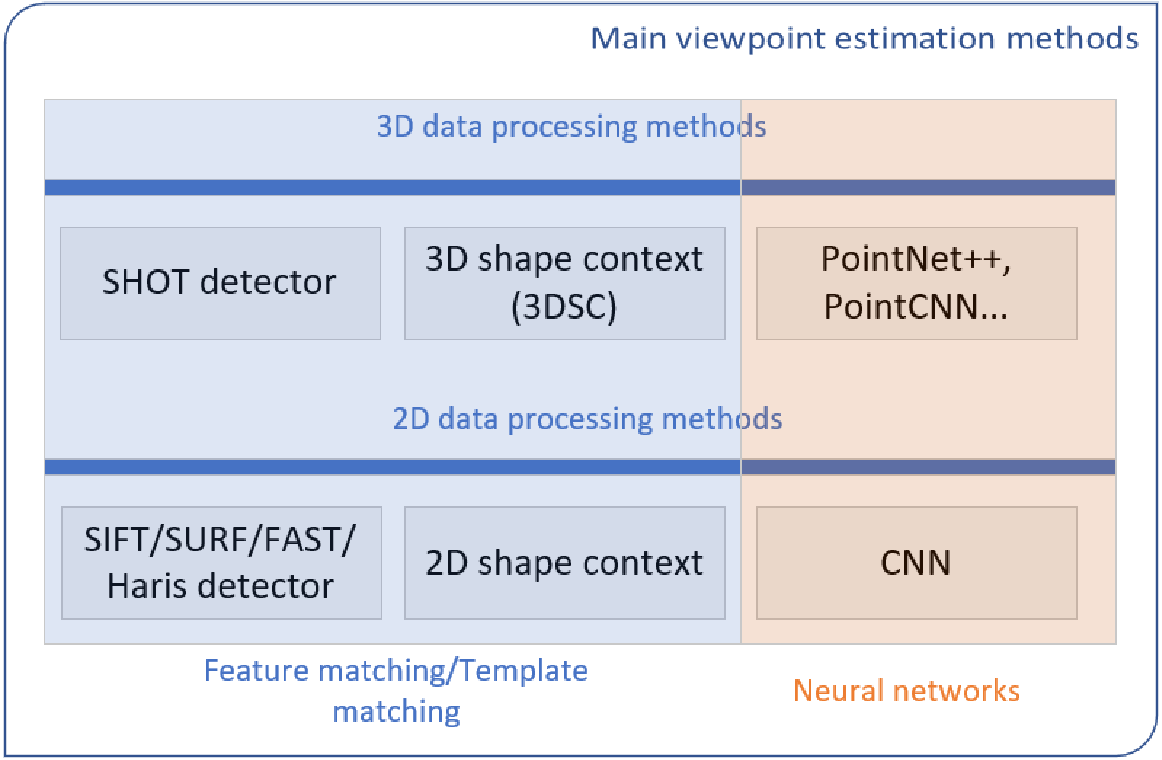
Introduction

- **Viewpoint estimation**
  - It is ***a fundamental step*** to further precisely compute the pose of target object, especially in the coarse-to-fine pose measure framework
  - Estimating the viewpoint of target object is an important step to robot manipulation, such as grasping
- **Examples**

Existing methods

- **Data type**
  - 3D data, such as point cloud, Computer-Aided Model (CAD)
  - 2D data, real or synthetic images

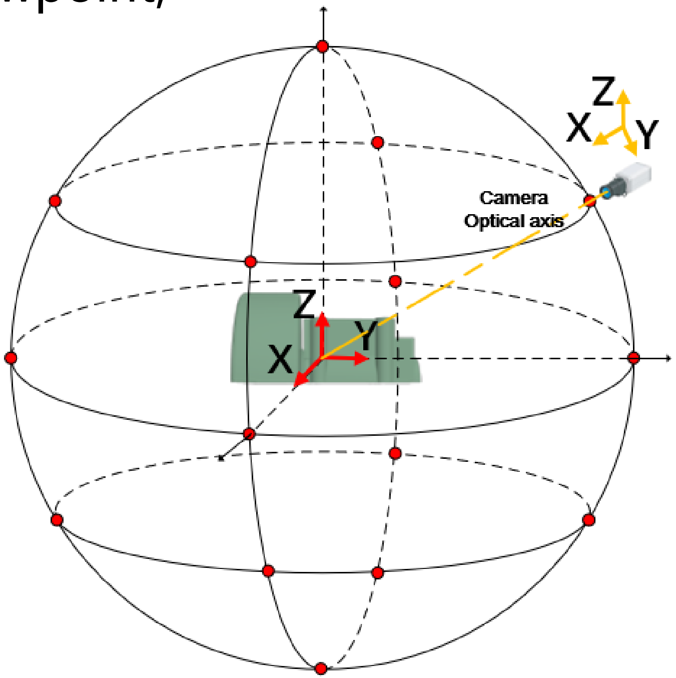
- **Existing viewpoint estimation methods**



Viewpoint estimation

- **Problem statement**
  - The complete expression of camera viewpoint, e.g. Euler angle ZYX format: (ψ, θ, φ)
  - Reduced format using the spherical coordinate frame: (α, β)
    α: the azimuthal angle (longitude)
    β: the polar angle (latitude)

    Hint: Here we reduce the one dimension by categorizing the views of rotating around the optical axis as a class. (α, β)==(ψ, θ)



  - Thus the viewpoint estimation problem has been transformed as viewpoint classification problem

Viewpoint estimation

- **Existing problems**
  - Traditional features are computationally heavy
  - The CNN based methods rely on huge annotated data
- **Inspirations**
  - Render CAD models to augment image datasets
  - Use transfer learning to bridge the gap between real images and synthetic images
  - Expect to estimate viewpoints of real object with the deep learning model which is only trained with automatically labeled CAD model images and unlabeled real counterpart images
  - Expect the deep model trained with labeled CAD model images can also work in real environment (Future work)

Outline

- Introduction
- Deep transfer networks with cold-to-hot training strategy
- Experimental results
- Conclusion

Deep transfer networks with cold-to-hot training strategy

- **Notations**
  - Discretized viewpoint space: V
  - Viewpoint: v $(v \in V)$
  - Synthetic images with annotations, which are rendered from CAD models (Source Domain)

$$\mathcal{T}^s = \{x_i^s, y_i^s\}$$

  - Unlabeled real-world workpiece images (Target Domain)

$$\mathcal{T}^t = \{x_i^t\}$$

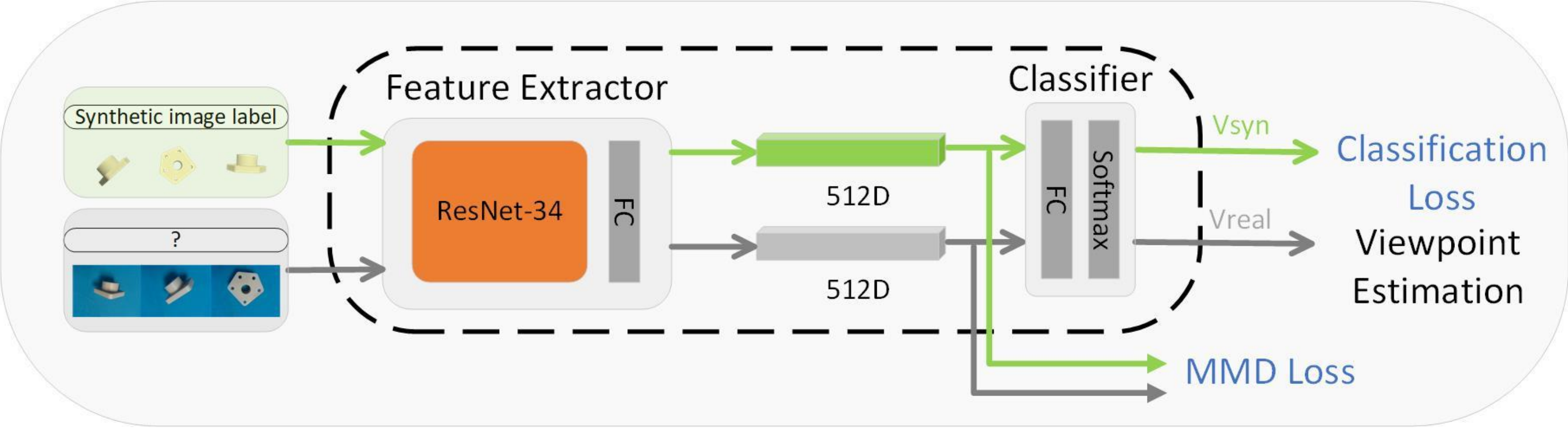
  - Training set

$$\mathcal{T} = \mathcal{T}^s \cup \mathcal{T}^t$$

Deep transfer networks with cold-to-hot training strategy

Deep transfer network (Basic version)

- Use ResNet-34 as feature extractor f
- Use multiple fully connected layers as classifier g



Deep transfer networks with cold-to-hot training strategy

Loss function

- Employ the multi-kernel MMD to align the high-level feature distributions between source domain and target domain. MMD loss

$$\mathcal{L}_{\mathcal{MMD}} = \left\| \frac{1}{|B^{s}|} \sum_{x_i^s \in B^s} \phi(f(x_i^s)) - \frac{1}{|B^t|} \sum_{x_j^t \in B^t} \phi(f(x_j^t)) \right\|_{\mathcal{H}}^2$$

where $\phi(\cdot) : \mathcal{X} \rightarrow \mathcal{H}$ and $|B^{\bullet}|$ refers to the number of samples in batch $B^{\bullet}$

- Geometric aware viewpoint classification loss

$$\mathcal{L}_{\mathcal{CLS}} = -\sum_{x_i^s \in B^s} \sum_{v \in \mathcal{V}} w(v, y_i^s) y_i^s log(P_v(x_i^s))$$

- Joint loss function

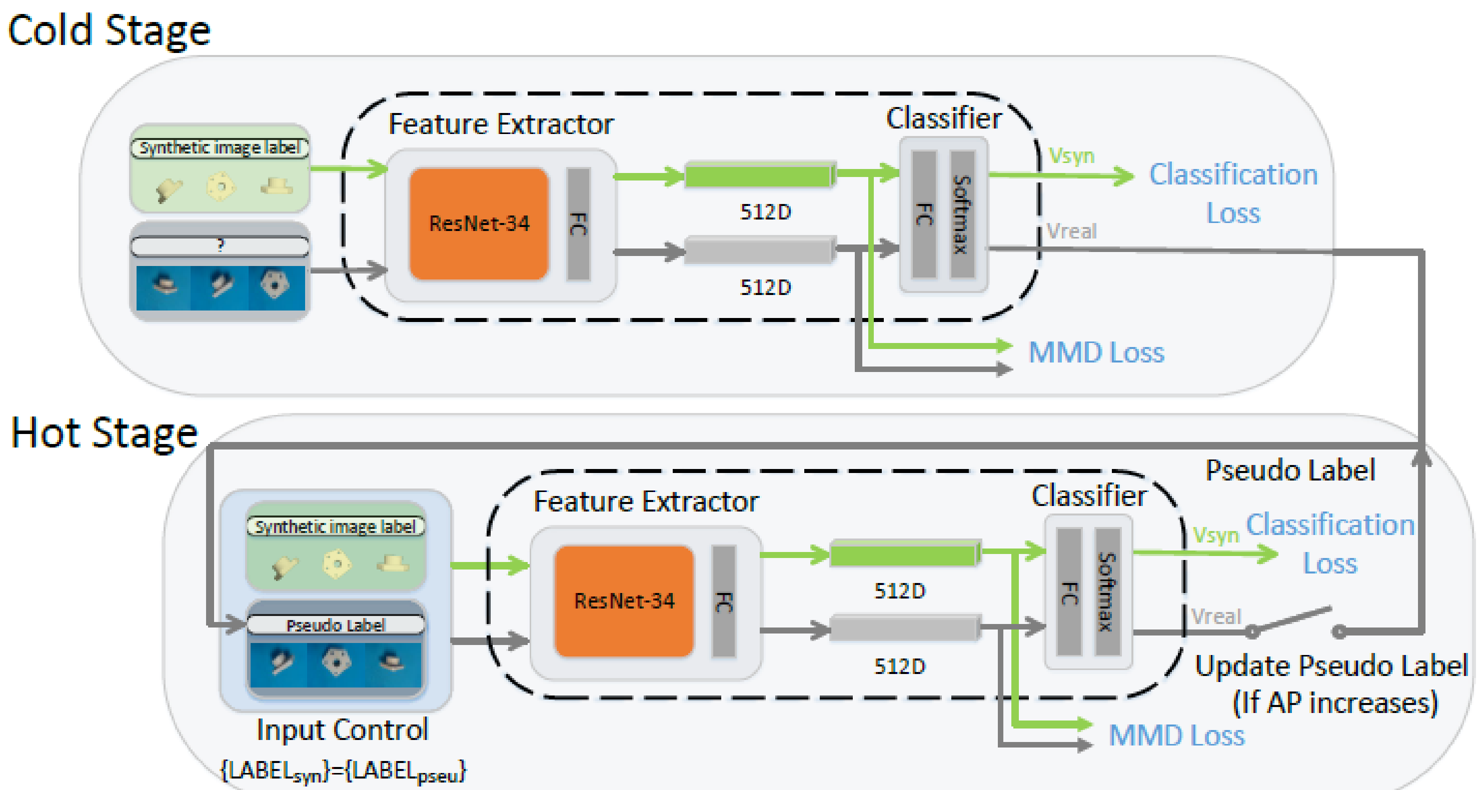
$$\mathcal{L}(\theta_f, \theta_g; \mathcal{T}) = \mathcal{L}_{\mathcal{CLS}} + \lambda \mathcal{L}_{\mathcal{MMD}}$$

Question:
If the real workpiece's viewpoint classification accuracy has reached 65%, how to get higher?

Deep transfer networks with cold-to-hot training strategy

- **Deep transfer networks with cold-to-hot training**
  - Feedback the pseudo labels of real workpiece
  - Use these pseudo labels for input distribution control.

    Namely let the input distribution of synthetic images to be same with that of real images



Deep transfer networks with cold-to-hot training strategy

- **Concrete implementation**
  - With pseudo labels, the real image set can be reformulated as

$$\tilde{\mathcal{T}}^t = \{x_i^t, \tilde{y}_i^t\}$$

  - Randomly select $|B^s|P^s(v)$ real image samples from set

$$\{x_i^t | (x_i^t, \tilde{y}_i^t) \in \tilde{T}^t, \tilde{y}_i^t = v\}$$

    to form the input batch of real images $\tilde{B}^t$

  - The difference of distributions of input batches between cold training stage and hot training stage is

$$\text{cold}\begin{cases} B^s \sim P^s(v) \\ B^t \sim P^t(v) \\ P^s(v) \neq P^t(v) \end{cases} \Rightarrow \text{hot}\begin{cases} B^s \sim P^s(v) \\ \tilde{B}^t \sim \tilde{P}^t(v) \\ P^s(v) = \tilde{P}^t(v) \end{cases}$$

Deep transfer networks with cold-to-hot training strategy

Reasons behind input distribution control

- The deep transfer network should also work when inputting two identical distribution batches B^s and $\tilde{B}^t$ from source domain and target domain. because the expectancy of closer distance between $f(Bs)$ and $f(\tilde{B}^t)$ is reasonable from the perspective of distance measure of MMD

- We believe that the deep model could learn more essential features and increase transfer ability if networks are fed with the data that are with higher correlation

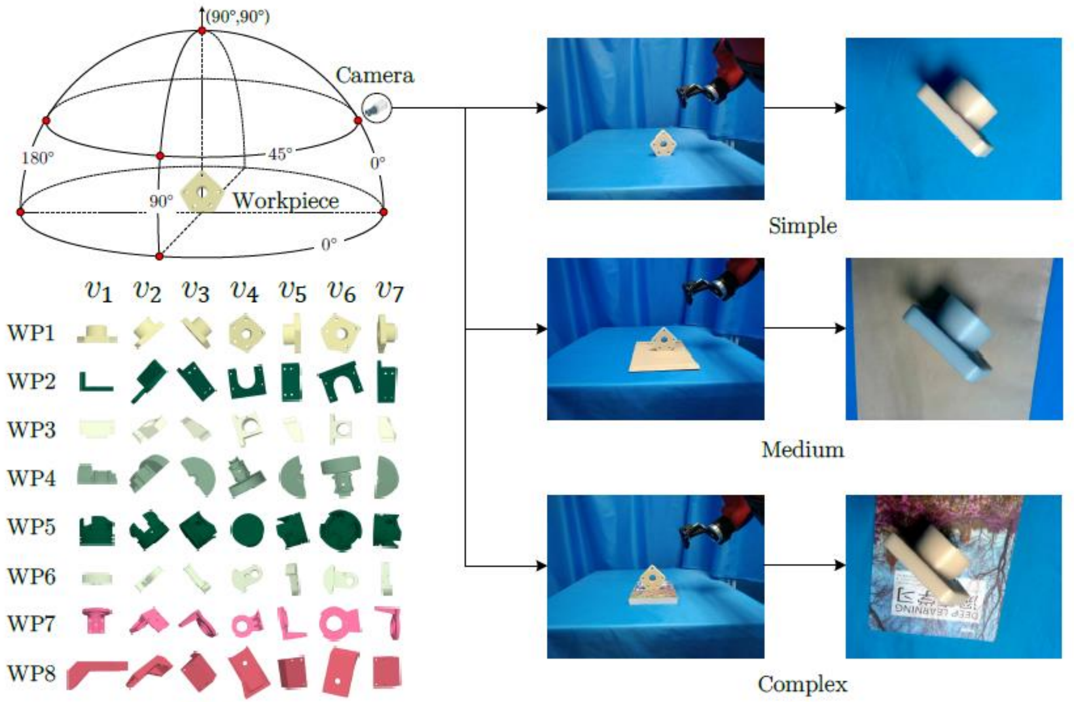
Outline

- Introduction
- Deep transfer networks with cold-to-hot training strategy
- **Experimental results**
- Conclusion

Experimental results

Datasets

- 12,400 synthetic images and 840 real workpiece images
- For simplicity, the 7 frontal viewpoints (marked by red dots) in the upper hemisphere are chosen
  - 4 different longitudes (0º, 90º, 180º, 270º)
  - 3 different latitudes (0º, 45º, 90º)



(a) (b)

Experimental results

Compared methods

- DDC (Tzeng et al., Computer Science 2014)
- DAN (Long et al., ICML 2015)
- JAN (Long et al., ICML 2017)

Tzeng, E., Homan, J., Zhang, N., Saenko, K., Darrell, T.: Deep domain confusion: Maximizing for domain invariance. Computer Science (2014)

Long, M., Cao, Y., Wang, J., Jordan, M.: Learning transferable features with deep adaptation networks. In: International Conference on Machine Learning. pp. 97-105 (2015)

Long, M., Zhu, H., Wang, J., Jordan, M.I.: Deep transfer learning with joint adaptation networks. In: International Conference on Machine Learning. pp. 2208-2217 (2017)

Experimental results

- **Results**
  - (+) means applying the cold-to-hot training strategy

Table 1. Mean average precision (mAP) across all viewpoint classes from workpiece one (WP1) to workpiece eight (WP8). The last column is the average of mAPs.

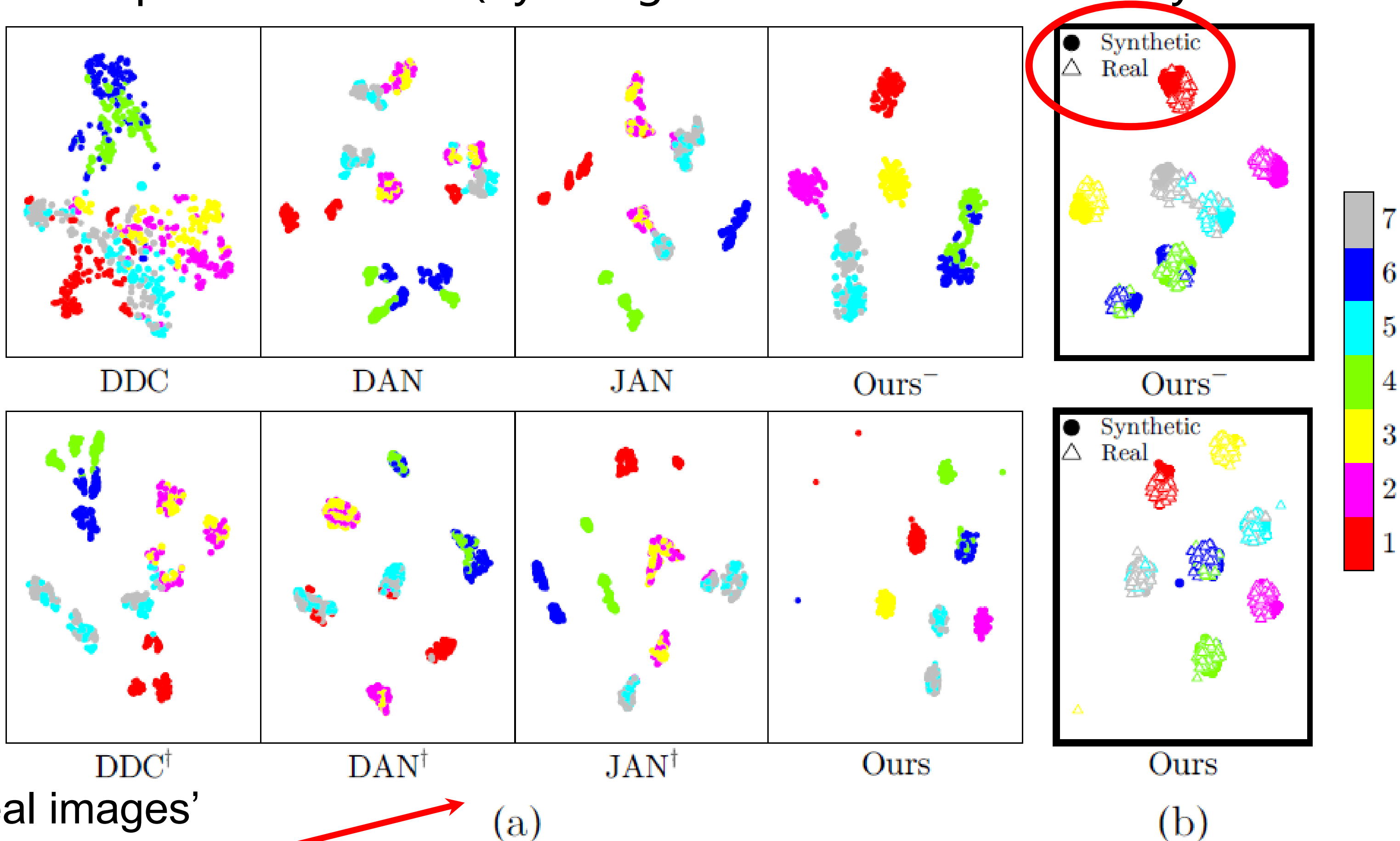
Method	WP1	WP2	WP3	WP4	WP5	WP6	WP7	WP8	Avg.
DDC	60.0%	46.6%	48.4%	48.3%	53.0%	53.2%	48.4%	79.8%	54.7%
DAN	70.9%	55.4%	49.5%	49.8%	70.2%	65.5%	70.2%	52.9%	60.6%
JAN	78.8%	61.5%	56.3%	66.6%	70.1%	66.0%	**80.9%**	73.2%	60.2%
Ours$^-$	87.8%	79.9%	**59.0%**	72.2%	73.6%	67.7%	73.6%	**88.8%**	75.3%
Ours	**91.9%**	**87.4%**	57.8%	**77.5%**	**76.6%**	**69.9%**	74.0%	87.4%	**77.8%**

Table 2. Average precision (AP) of each viewpoint class on workpiece one (WP1).

Method	$(90°, 90°)$	$(45°, 180°)$	$(45°, 0°)$	$(0°, 90°)$	$(0°, 180°)$	$(45°, 90°)$	$(0°, 0°)$	mAP
DDC	82.4%	66.0%	**72.9%**	66.8%	41.4%	73.8%	46.0%	60.0%
DAN	98.6%	51.2%	48.6%	97.4%	**56.8%**	90.1%	51.4%	70.9%
JAN	**99.9%**	42.6%	54.9%	**100.0%**	**59.4%**	95.2%	44.1%	**78.8%**
DDC†	**92.5%**	**70.8%**	69.4%	**78.5%**	**69.9%**	**71.7%**	**57.1%**	**69.3%**
DAN†	**99.8%**	**51.3%**	**55.3%**	**99.5%**	51.6%	**96.6%**	**53.3%**	**77.1%**
JAN†	**99.9%**	**52.8%**	**58.2%**	99.9%	49.9%	**99.3%**	**48.9%**	76.9%
Ours$^-$	**100.0%**	99.7%	99.9%	89.5%	**84.7%**	77.8%	**84.7%**	87.8%
Ours	**100.0%**	**100.0%**	**100.0%**	**94.6%**	83.4%	**80.9%**	79.9%	**91.9%**

Experimental results

- **Visualization**
  - Visualization of the learned high-level features distributions of compared methods (by using t-SNE for dimensionality reduction)



The test real images' features distribution

Outline

- Introduction
- Deep transfer networks with cold-to-hot training strategy
- Experimental results
- Conclusion

Conclusion

- We propose a deep transfer network integrated with transfer ability, geometric aware loss and cold-to-hot training strategy for workpiece viewpoint estimation

- From the **large automatically labeled synthetic images** rendered by CAD models, the network can learn transfer the knowledge for estimating the viewpoint of **unlabeled real image**

- From beginning to end, the training set of deep transfer network is **without** the labels of real images, which is promising to evade manual work of annotation

Thank you!