

Semantically Meaningful View Selection



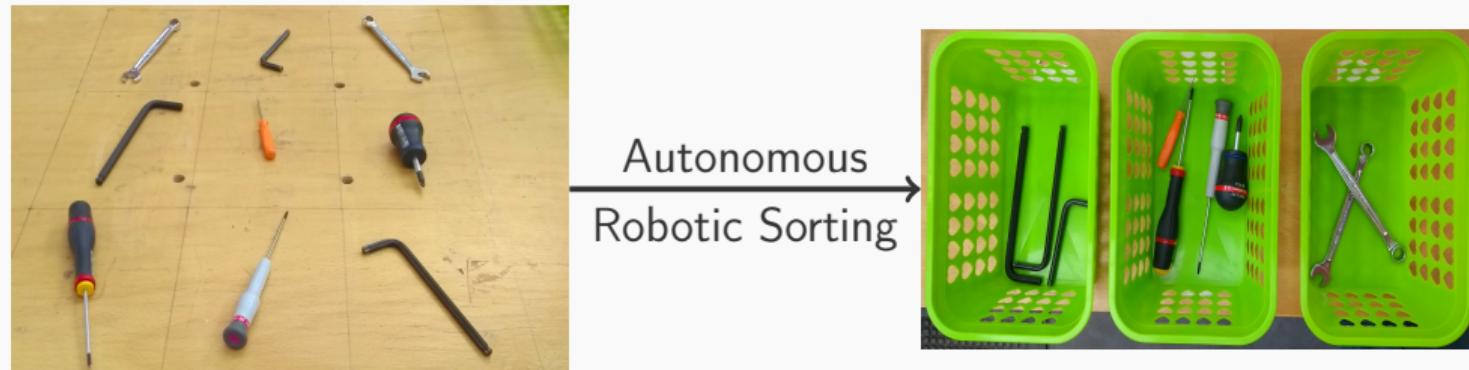
Joris Guérin¹ Olivier Gibaru¹
Eric Nyiri¹ Stéphane Thiery¹ Byron Boots²

¹ Arts et Métiers ParisTech, ² Georgia Institute of Technology

Motivations

Semantic object understanding is an important task

Example:



Importance of View Selection



(a) Top view

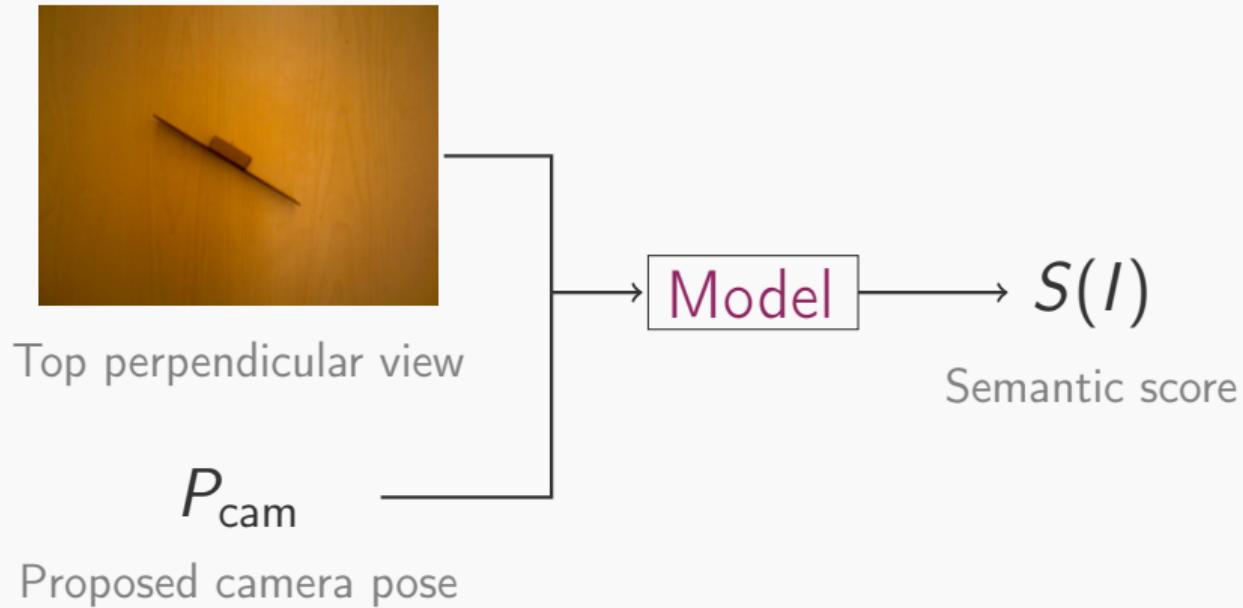


(b) Good view



(c) Bad view

Objective



Approach

- ▶ Build a **large multi-view dataset**

# Classes	# Object/class (<i>total</i>)	# Poses/object (<i>total</i>)	# Images/pose (<i>total</i>)
29	4-6 (<i>144</i>)	3 (<i>432</i>)	17-22 (<i>9112</i>)

- ▶ Fit a **custom clusterability score** to each view
- ▶ Train a CNN to **predict the score** for a given camera pose

Clusterability score

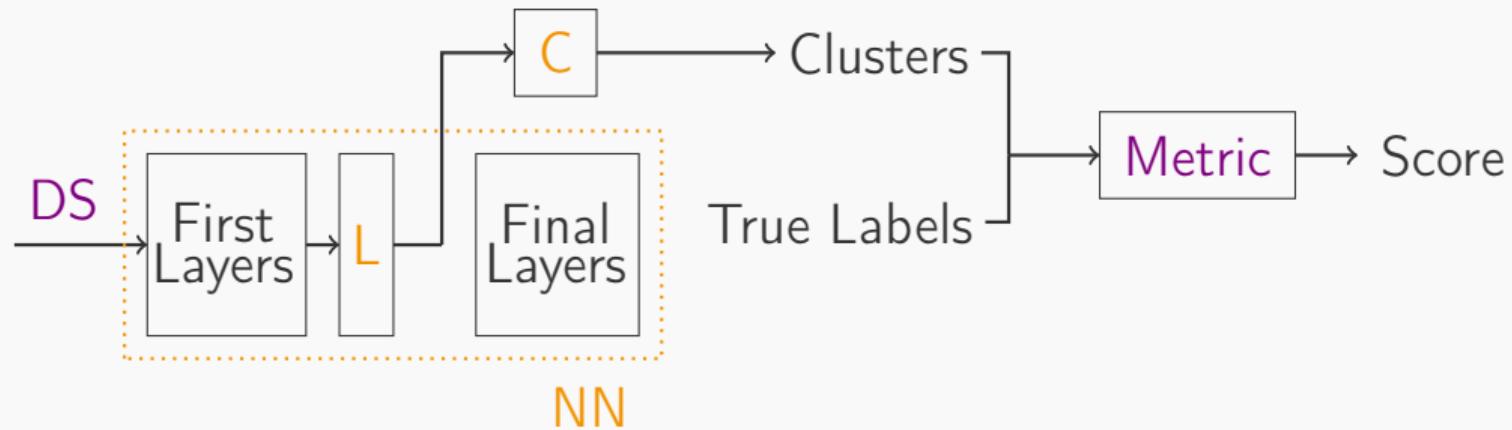
$$\text{CP}_{MC} = \{\text{cp}_i, i \in \{1, \dots, N\}\}$$

$$FMI_{\text{cp}, a}^i = \frac{TP_i}{\sqrt{(TP_i+FP_i)(TP_i+FN_i)}} \quad \text{CP}_{MC} \text{ such that}$$

$$\tilde{s}_{\text{cp}}(I) = \begin{cases} FMI_{\text{cp}}^I & \text{if } I \in \text{cp} \\ 0 & \text{otherwise} \end{cases} \quad \min(\{N^I, I \in X\}) \geq 2 \times 10^5$$
$$N \approx 3 \times 10^7$$

$$\hat{s}(I) = \sum_{\text{cp} \in \text{CP}_{MC}} \tilde{s}_{\text{cp}}(I) / N^I$$

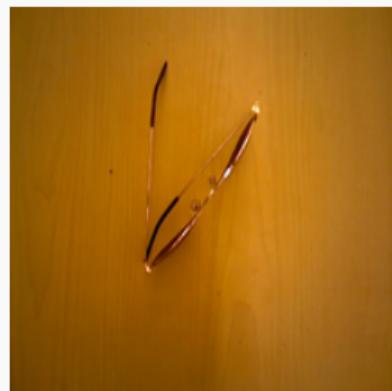
Clustering pipeline



NN = Xception, L = average pooling, C = agglomerative clustering

Approach

Example images and qualitative evaluation of clusterability index



Top view



(45, 45)

(45, 225)

(60, 45)

(60, 225)

(75, 45)

(75, 225)



(45, 135)



(45, 315)



(60, 135)



(60, 315)

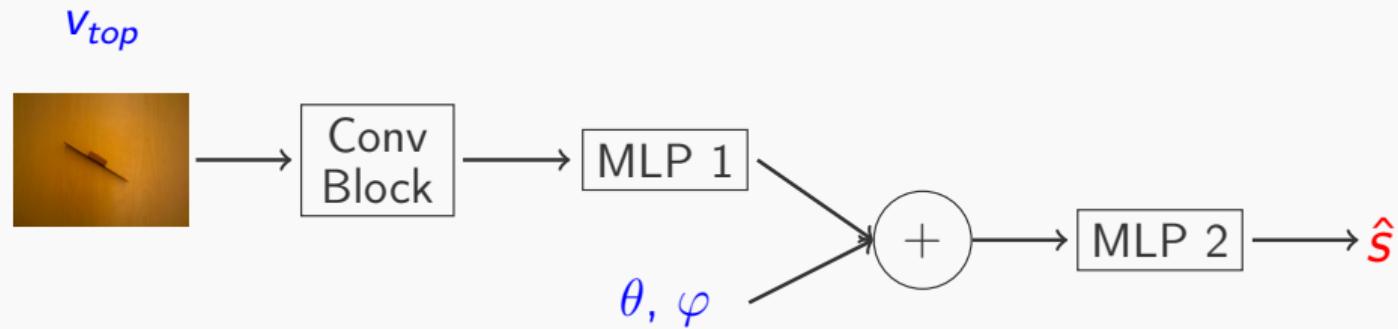


(75, 135)



(75, 315)

SVnet architecture



Qualitative evaluation of view selection network



(a) Example top views



(b) Associated SV-net selections

Quantitative results

Scores with top view, random view and view predicted by the network

	FM	NMI	PUR
TOP	0.44	0.51	0.70
RAND	0.48	0.56	0.74
SV-net	0.55	0.63	0.78

Averaged over 10^4 clustering problems