



BOSCH
Invented for life

Viewpoint Selection for Visual Failure Detection

Akanksha Saran¹, Branka Lakic², Srinoy Majumdar¹, Juergen Hess³, Scott Niekum¹

¹University of Texas at Austin

²Duke University

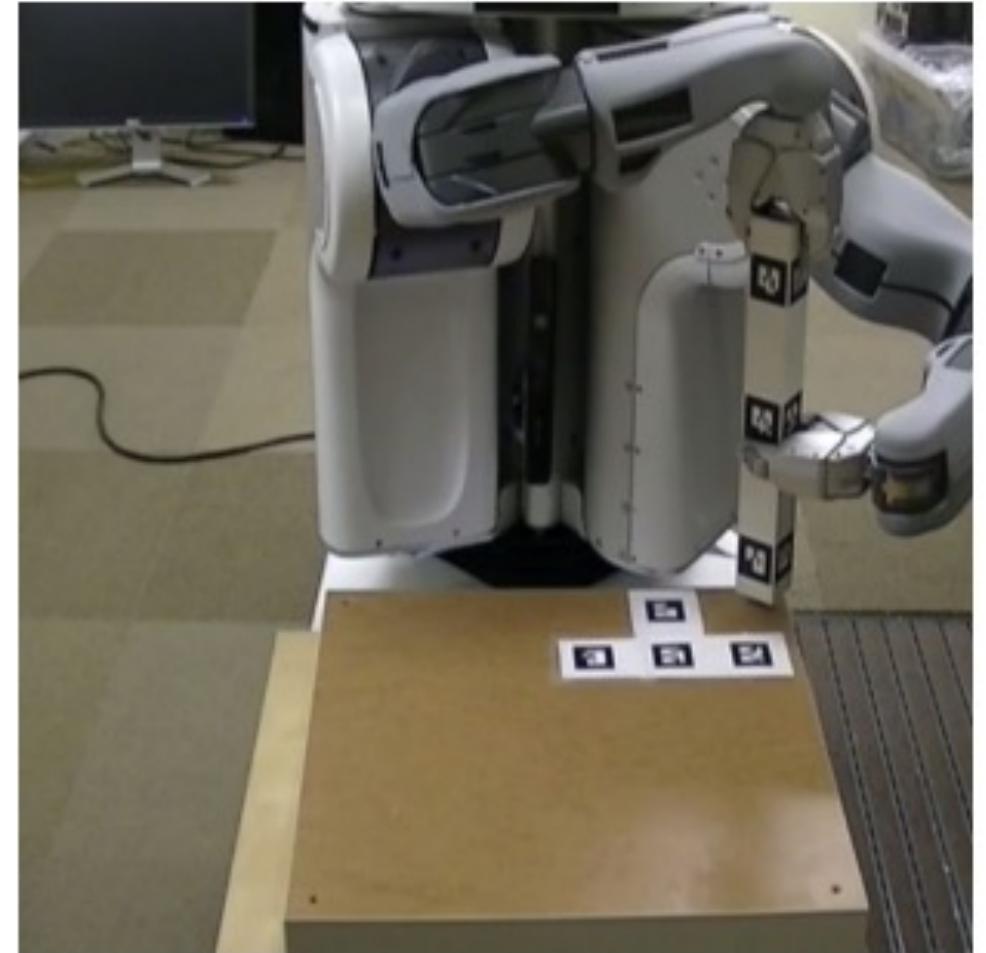
³Bosch Research and Technology Center

Why Failure Detection?

Why Failure Detection?



Why Failure Detection?



“Visual” Failure Detection

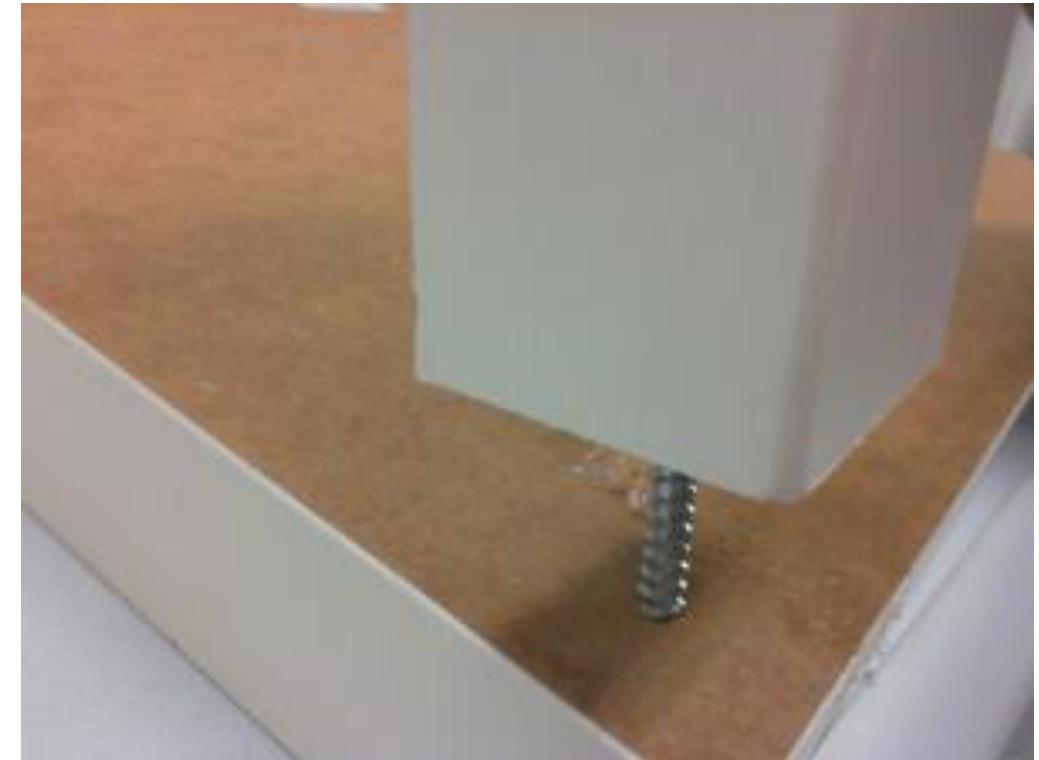


MICROSOFT

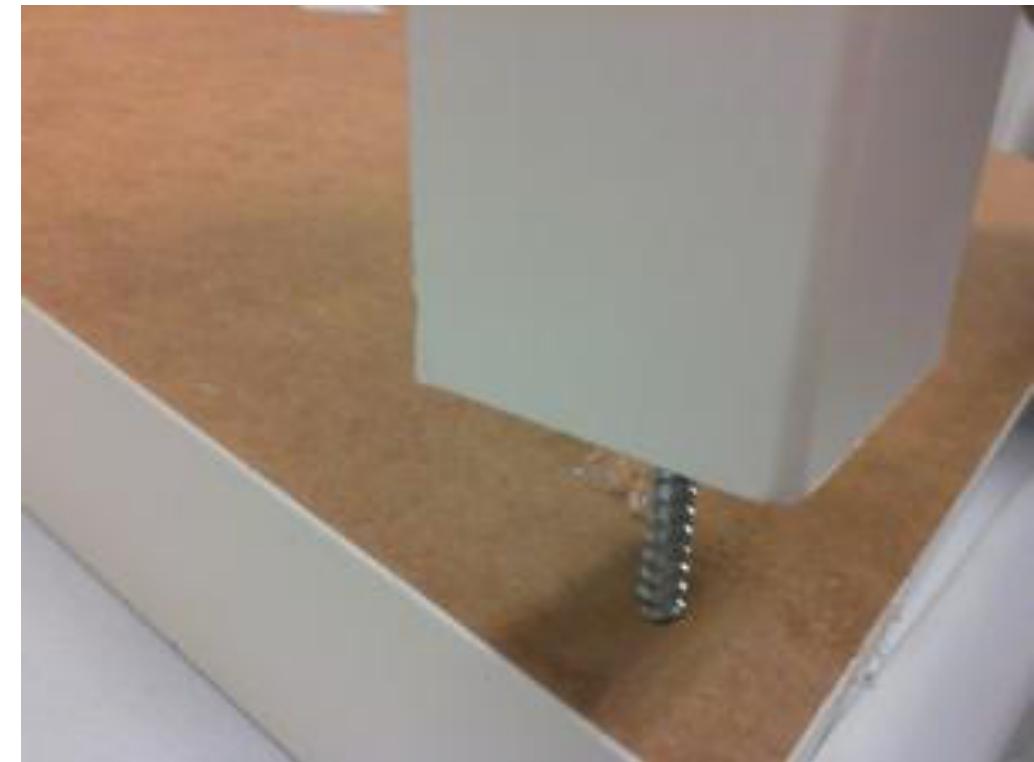


Challenges of Visual Failure Detection

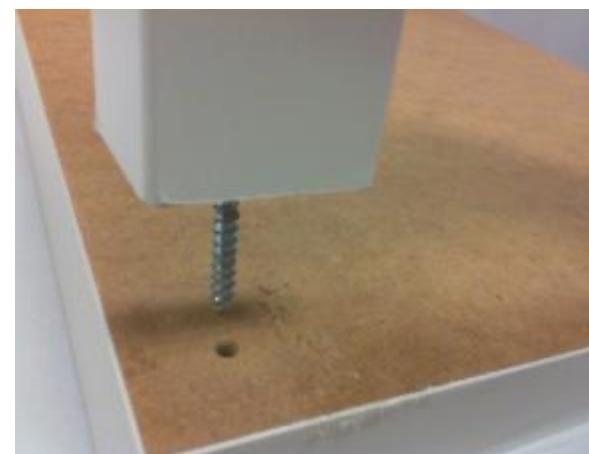
Only a “small portion” of the image contains
Useful Discriminative Information



Fine-Grained Visual Differences



Only Certain Viewpoints contain useful Discriminative Information



Approach

1. Discriminating between success and failure
2. Select a viewpoint from which failure detected most accurately

Approach

- 1. Discriminating between success and failure**
2. Select a viewpoint from which failure detected most accurately

Supervised Failure Detection Algorithms

1. Random forest-based method that discovers highly informative fine-grained visual features
2. SVM models trained on features extracted from pre-trained convolutional neural networks

Supervised Failure Detection Algorithms

- 1. Random forest-based method that discovers highly informative fine-grained visual features**
2. SVM models trained on features extracted from pre-trained convolutional neural networks

Fine-grained Random Forests



B. Yao*, A. Khosla*, and L. Fei-Fei, “Combining randomization and discrimination for fine-grained image categorization,” in IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2011.

Fine-grained Random Forests

Random Forests with Strong Classifiers



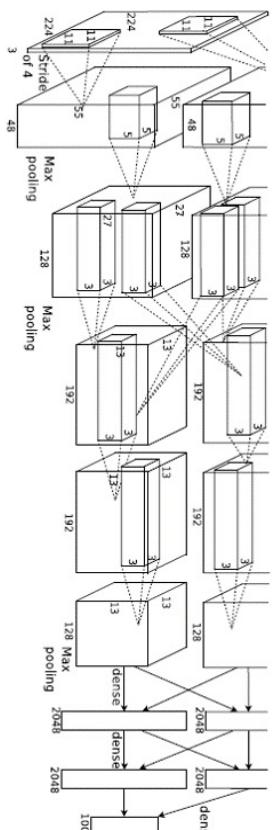
B. Yao*, A. Khosla*, and L. Fei-Fei, "Combining randomization and discrimination for fine-grained image categorization," in IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2011.

Supervised Failure Detection Algorithms

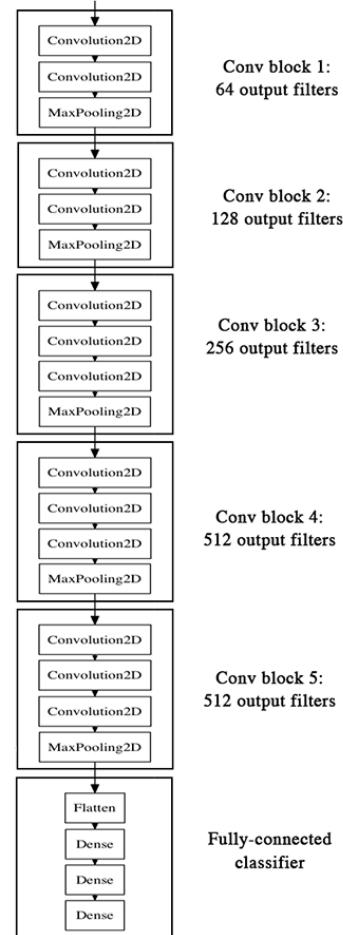
1. Random forest-based method that discovers highly informative fine-grained visual features
2. **SVM models trained on features extracted from pre-trained convolutional neural networks**

Deep Features + SVM

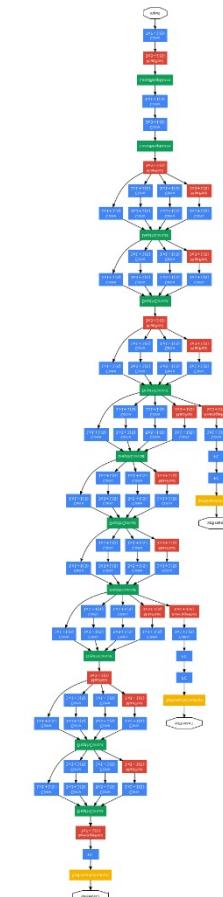
AlexNet



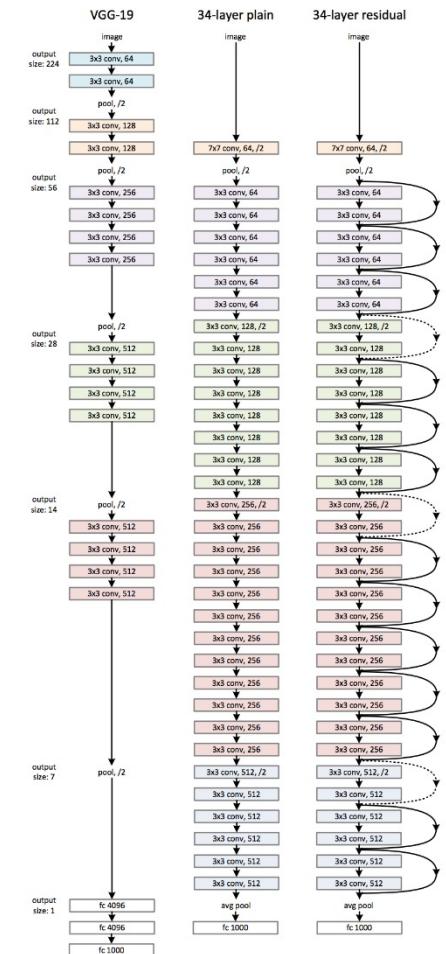
VGGNet



GoogLeNet



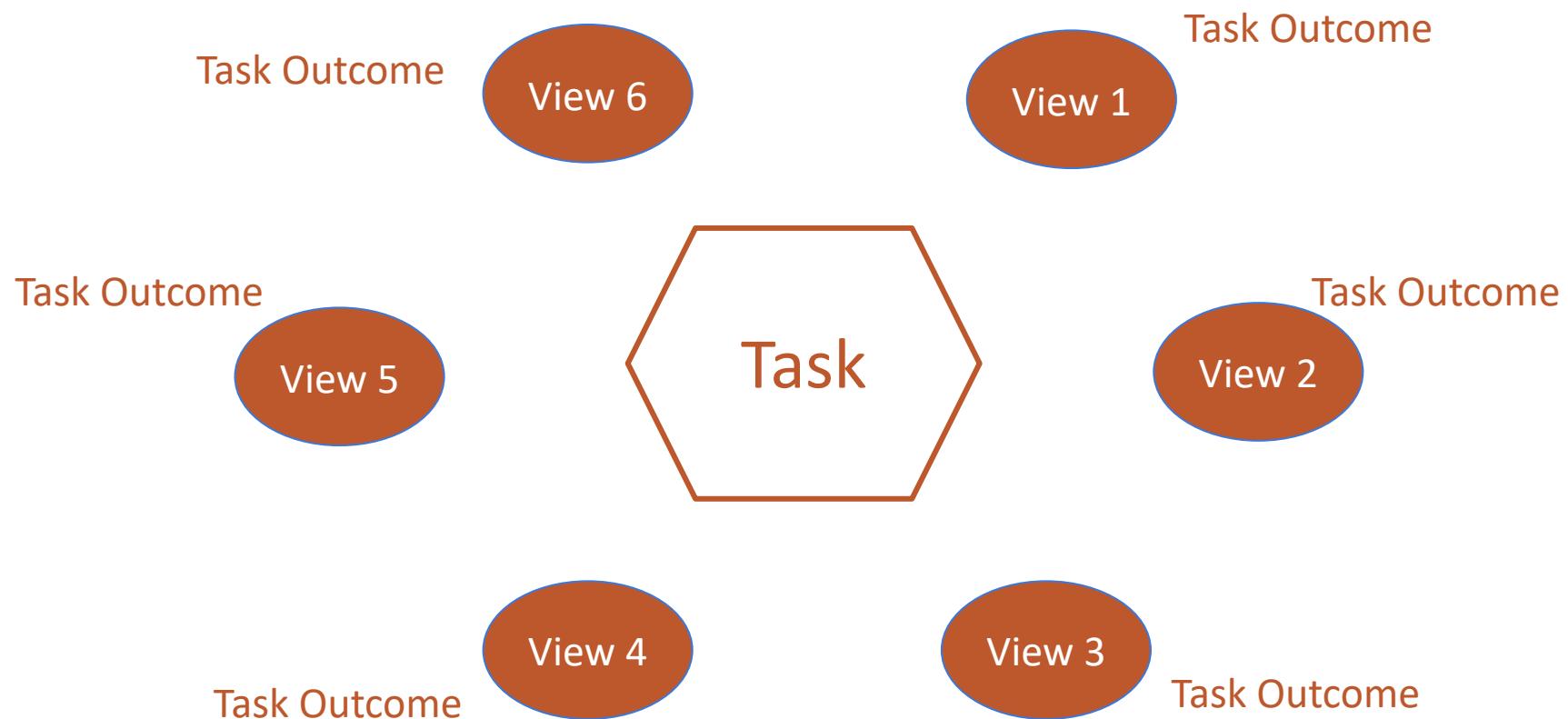
ResNet



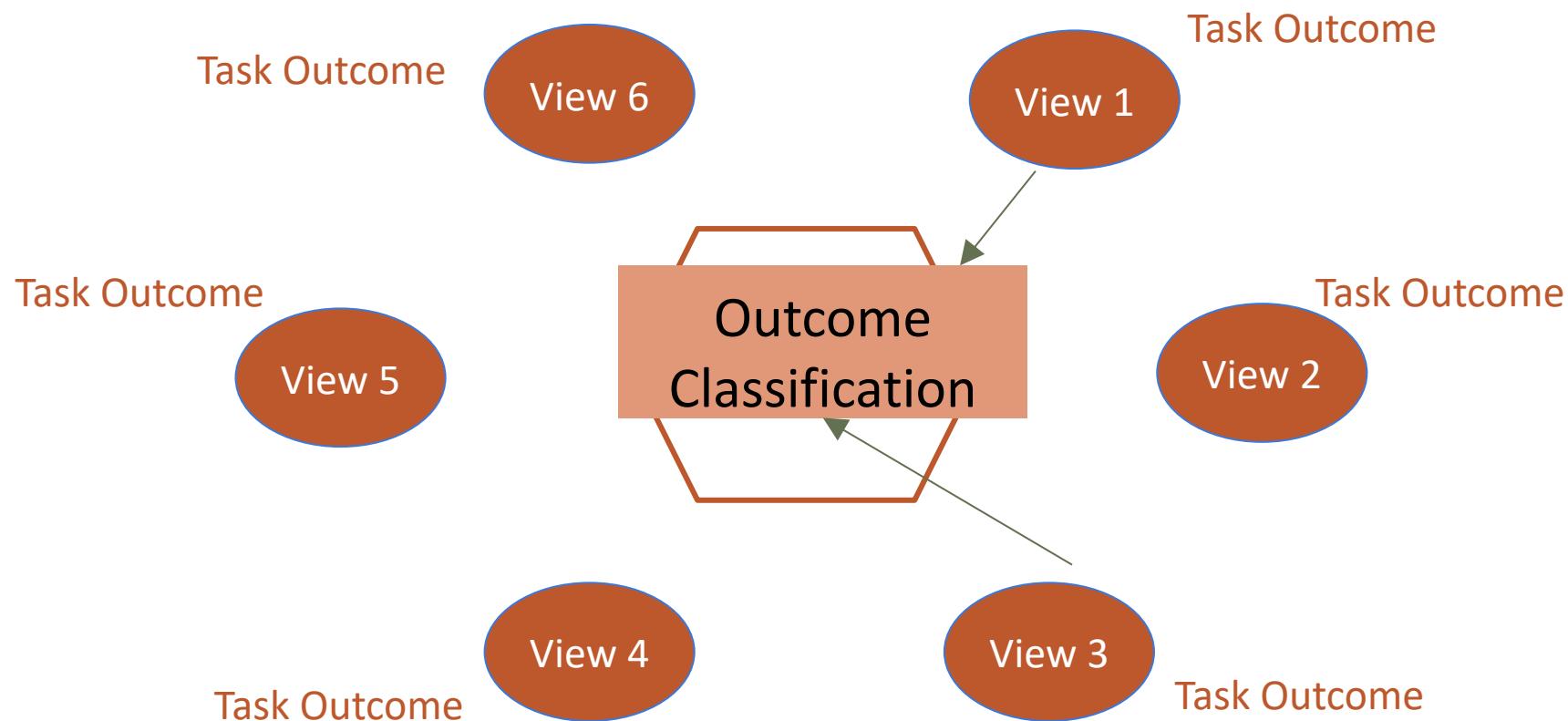
Approach

1. Discriminating between success and failure
2. Select a viewpoint from which failure detected most accurately

Viewpoint Selection

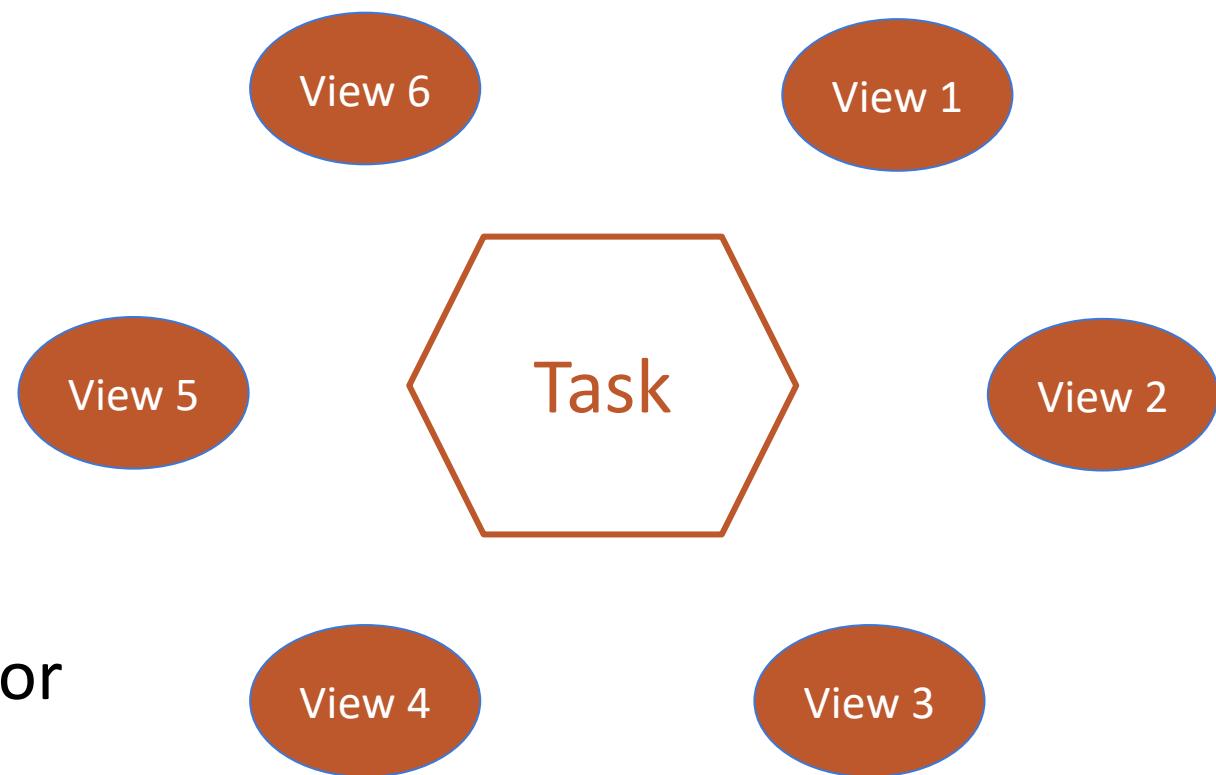


Viewpoint Selection

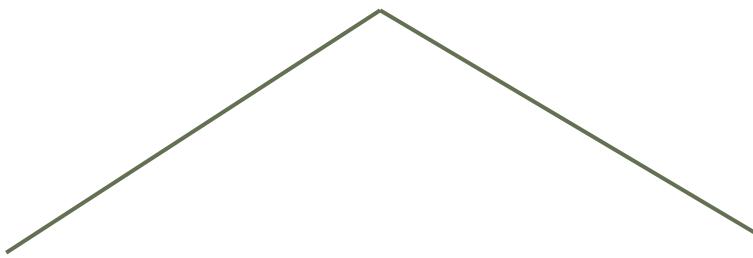


Our Assumptions for Viewpoint Selection

1. Viewpoints are at fixed locations with respect to task setup
2. Viewpoint selection and subsequent failure detection occurs at the end of task execution.
3. Object(s) of interest captured in one or more of the viewpoints.



Viewpoint Selection



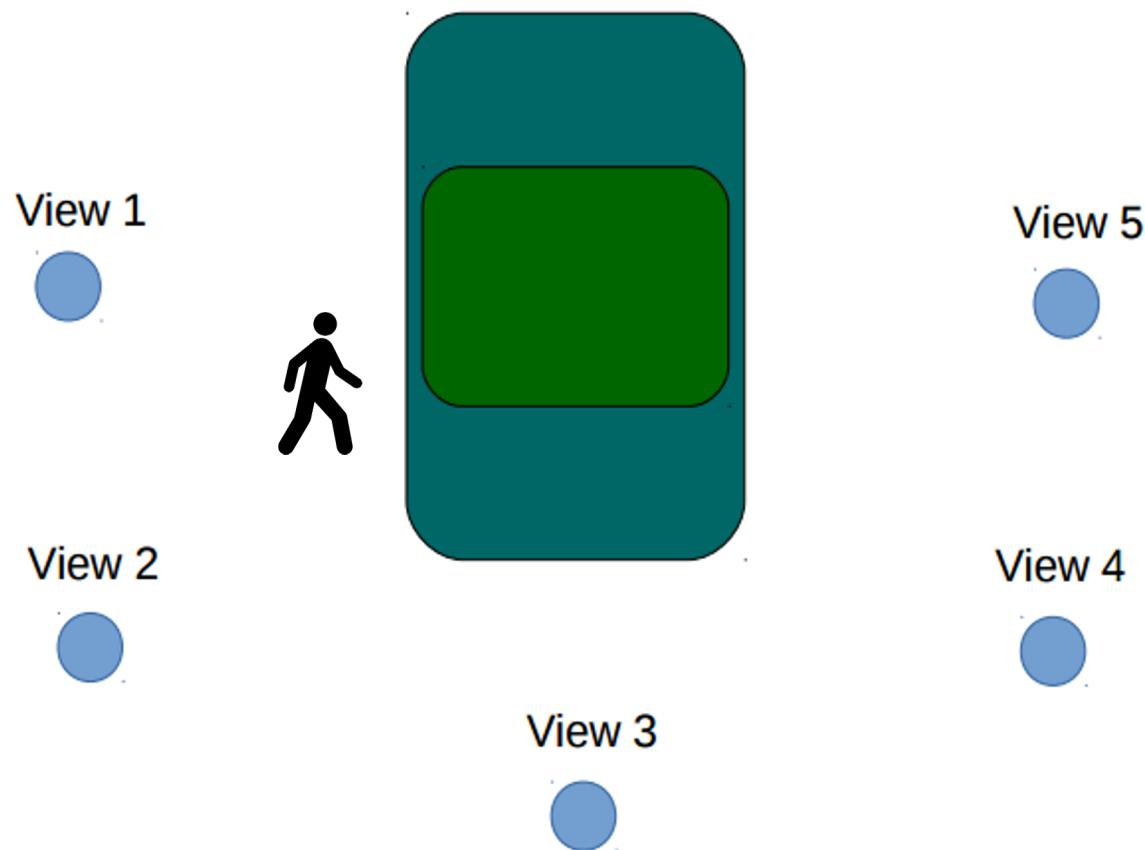
**Static Viewpoint
Selection**

Independent of the
features of the test image

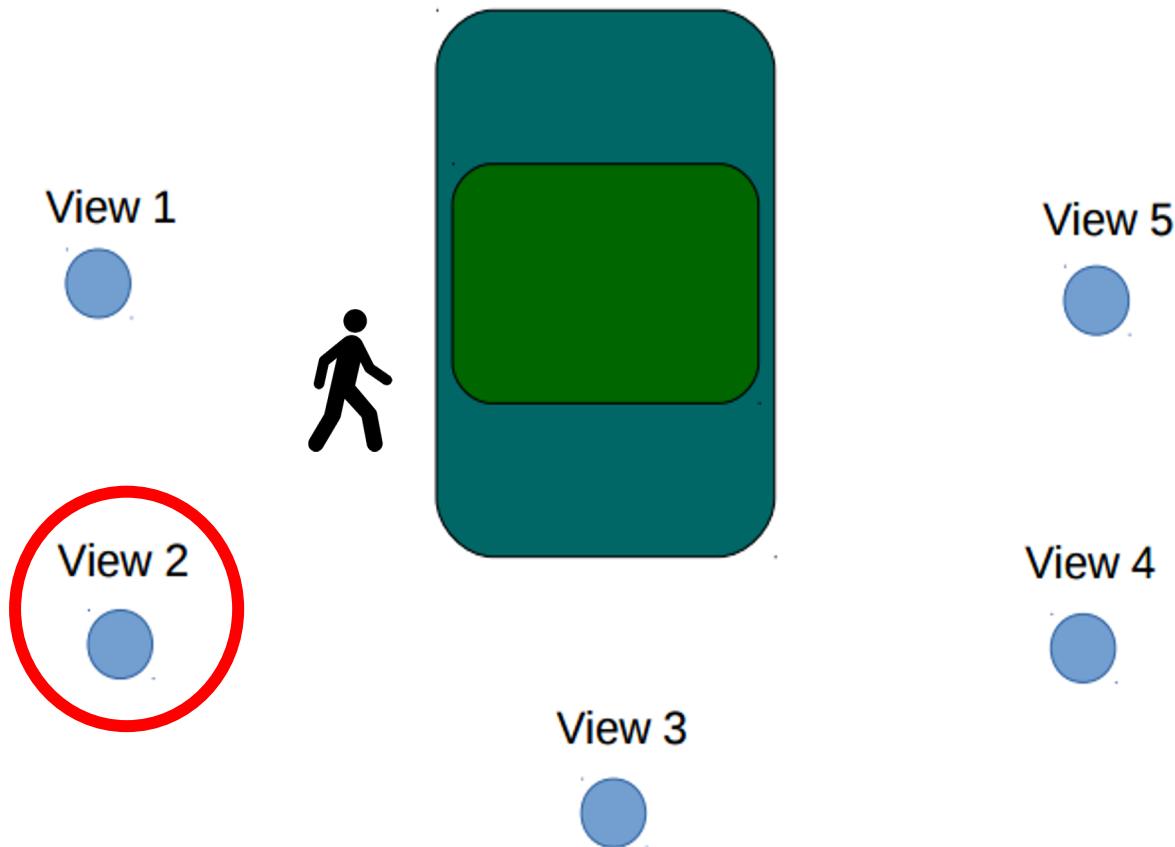
**Active Viewpoint
Selection**

Dependent on the
features of the test image

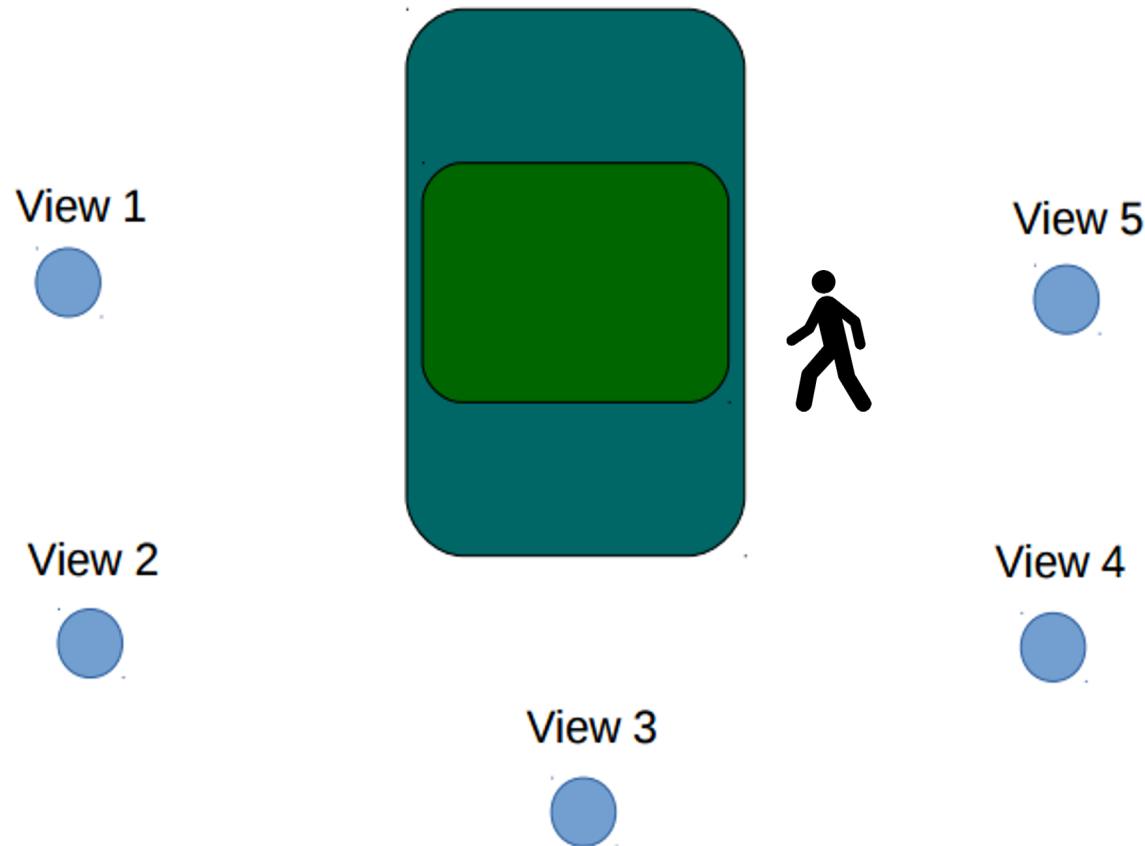
Static Viewpoint Selection



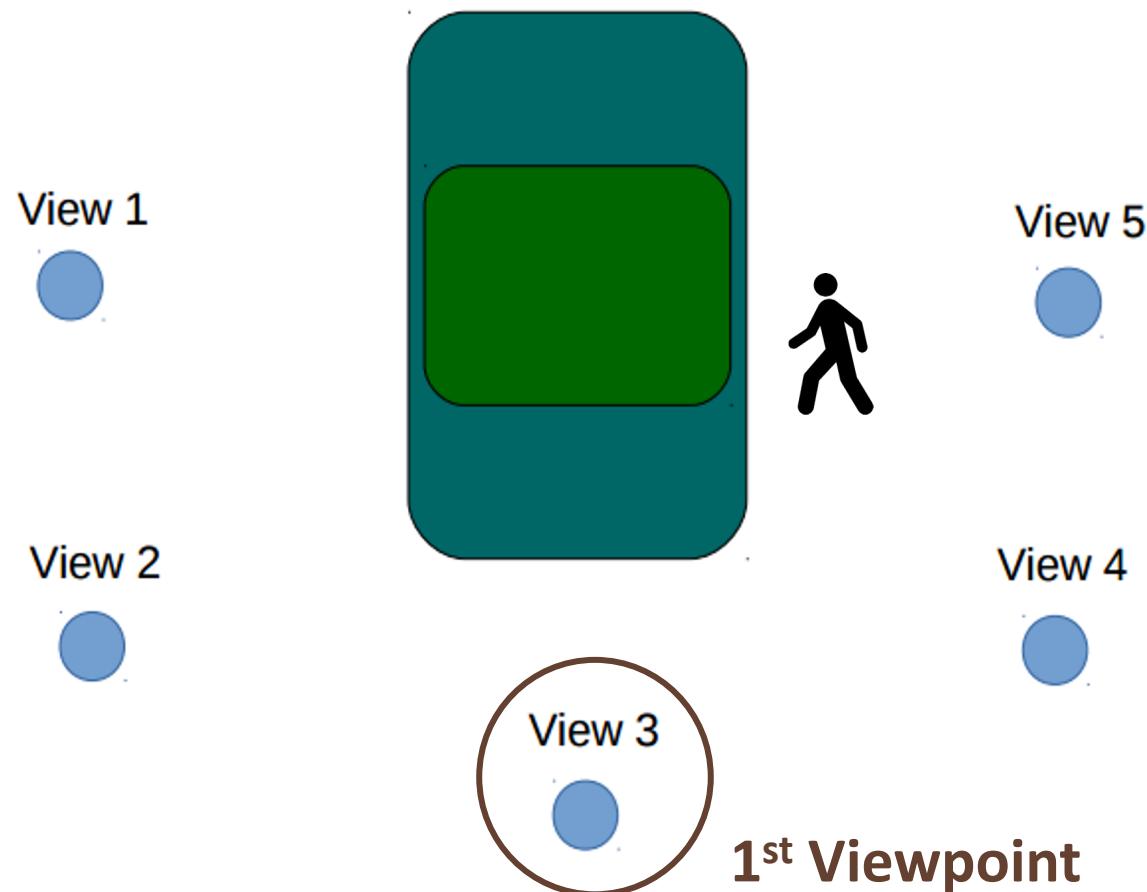
Static Viewpoint Selection



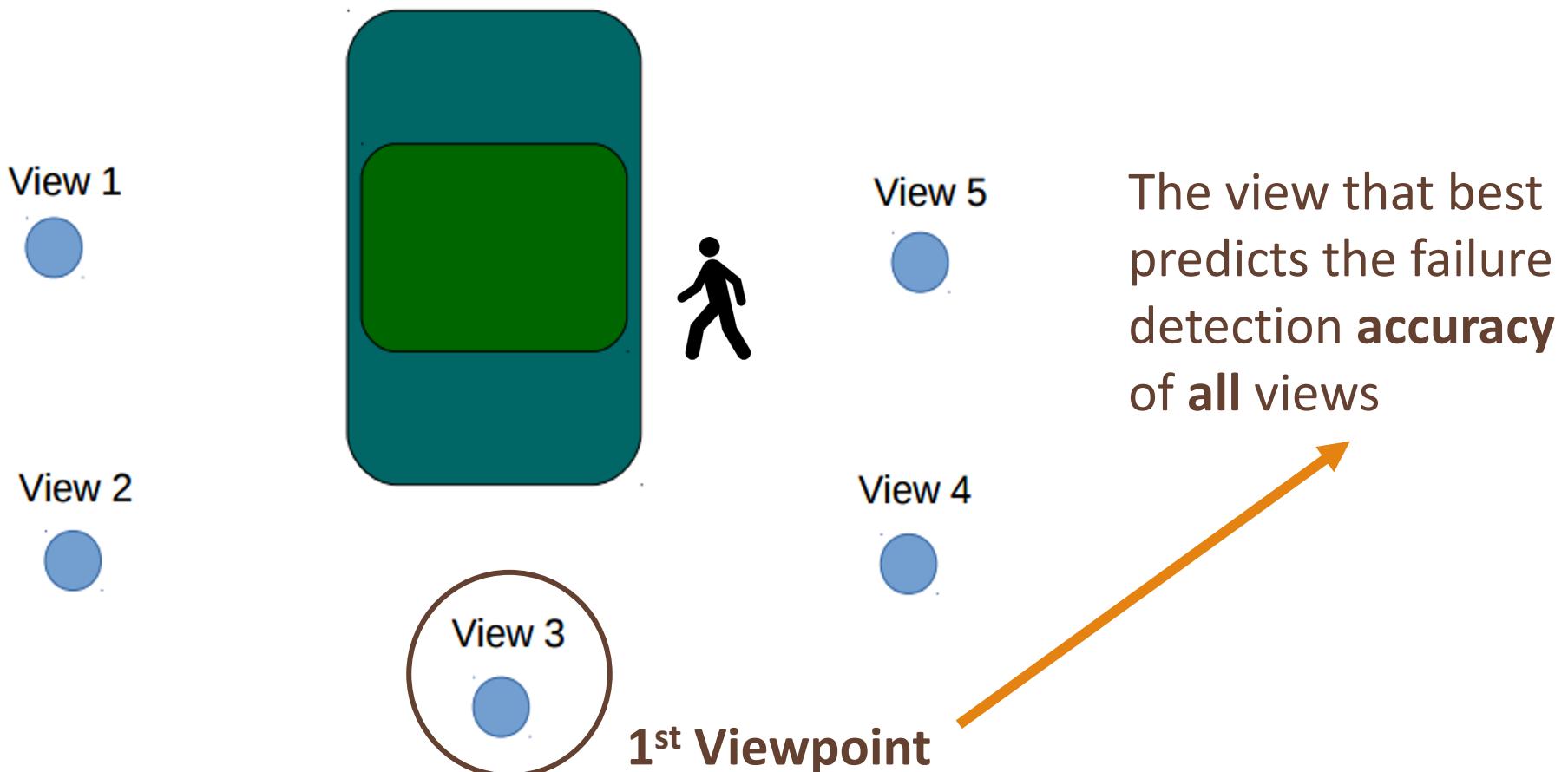
Active Viewpoint Selection



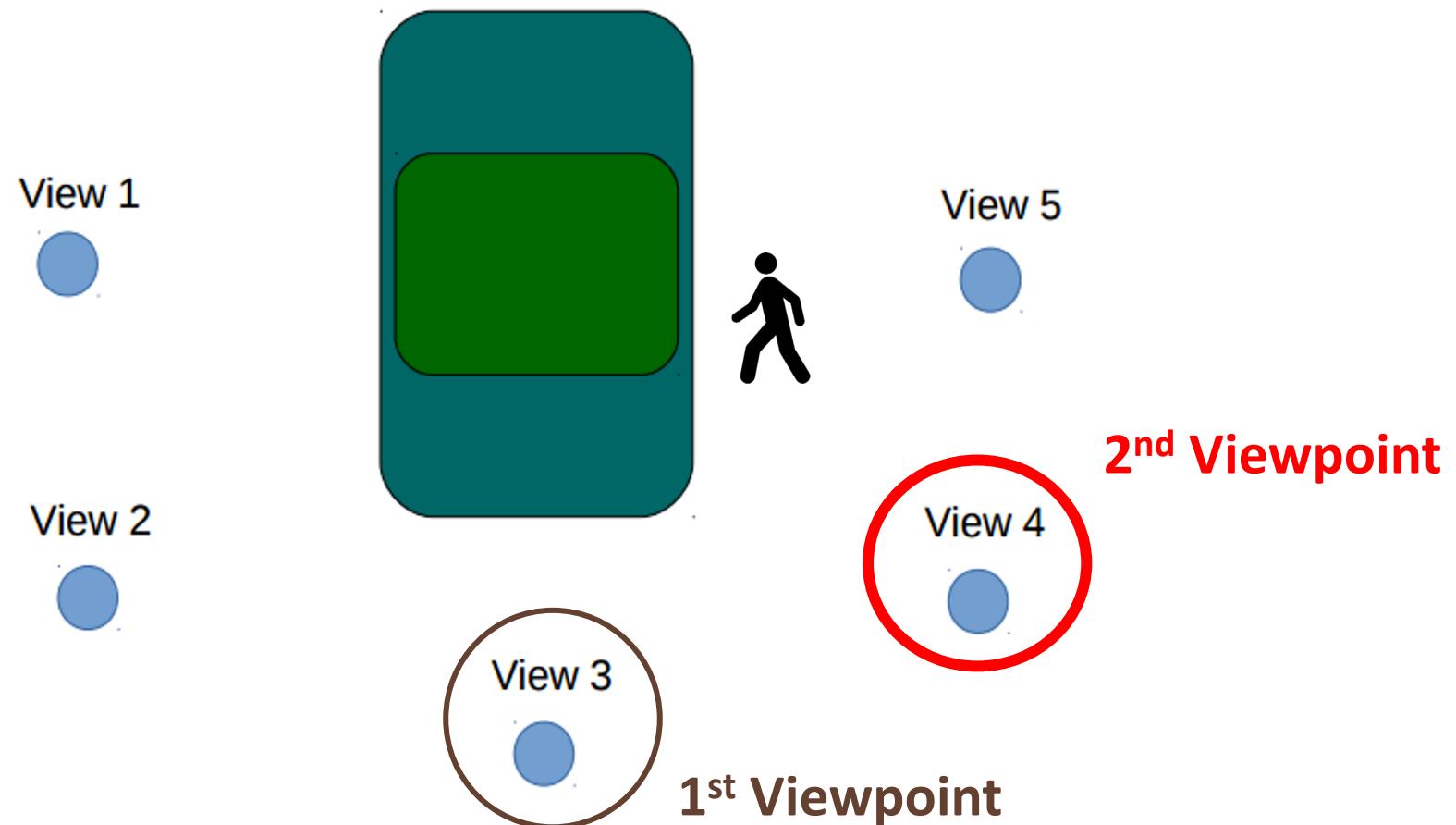
Active Viewpoint Selection



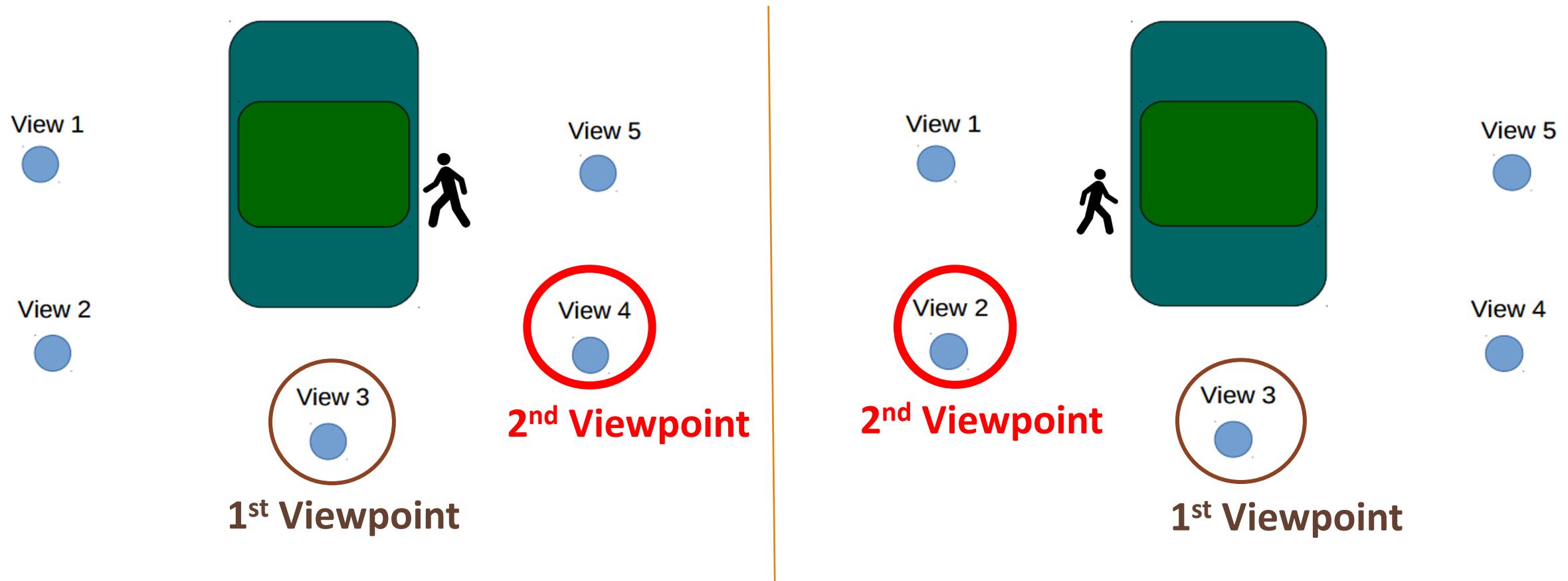
Active Viewpoint Selection



Active Viewpoint Selection



Active Viewpoint Selection



Experiments & Results

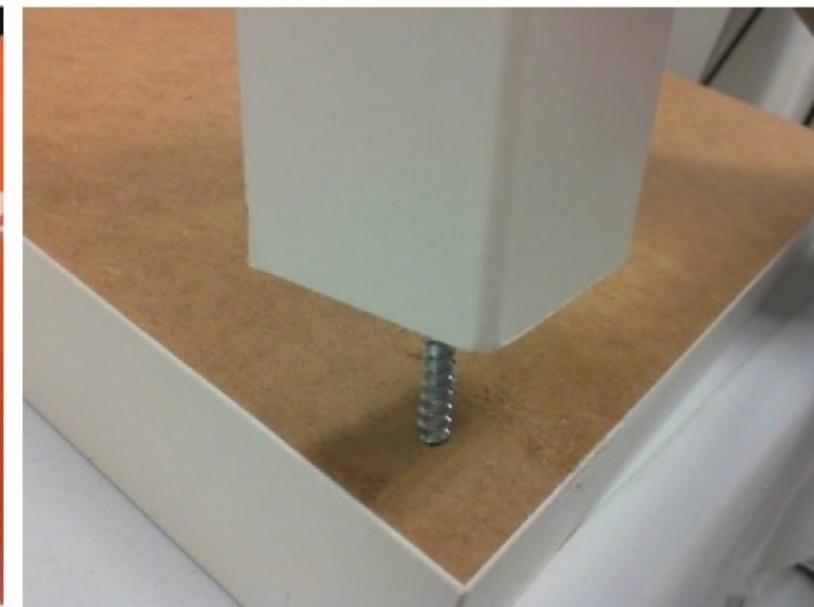
Experimental Domains

1. Ikea Table Assembly Task
 - a. Without Obstacles
 - b. With Dynamic Obstacles
2. Quadrotor Surveillance Task with Dynamic Human Location

Experimental Domains

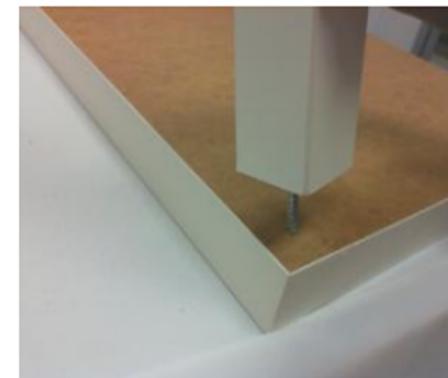
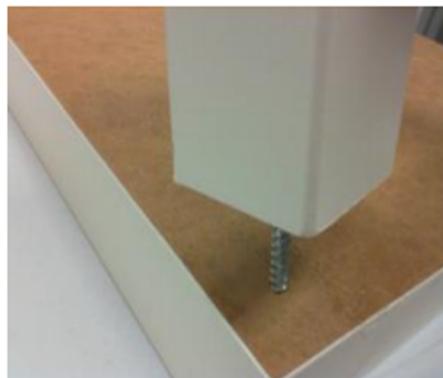
- 1. Ikea Table Assembly Task**
 - a. Without Obstacles**
 - b. With Dynamic Obstacles**
- 2. Quadrotor Surveillance Task with Dynamic Human Location**

Ikea Table Assembly

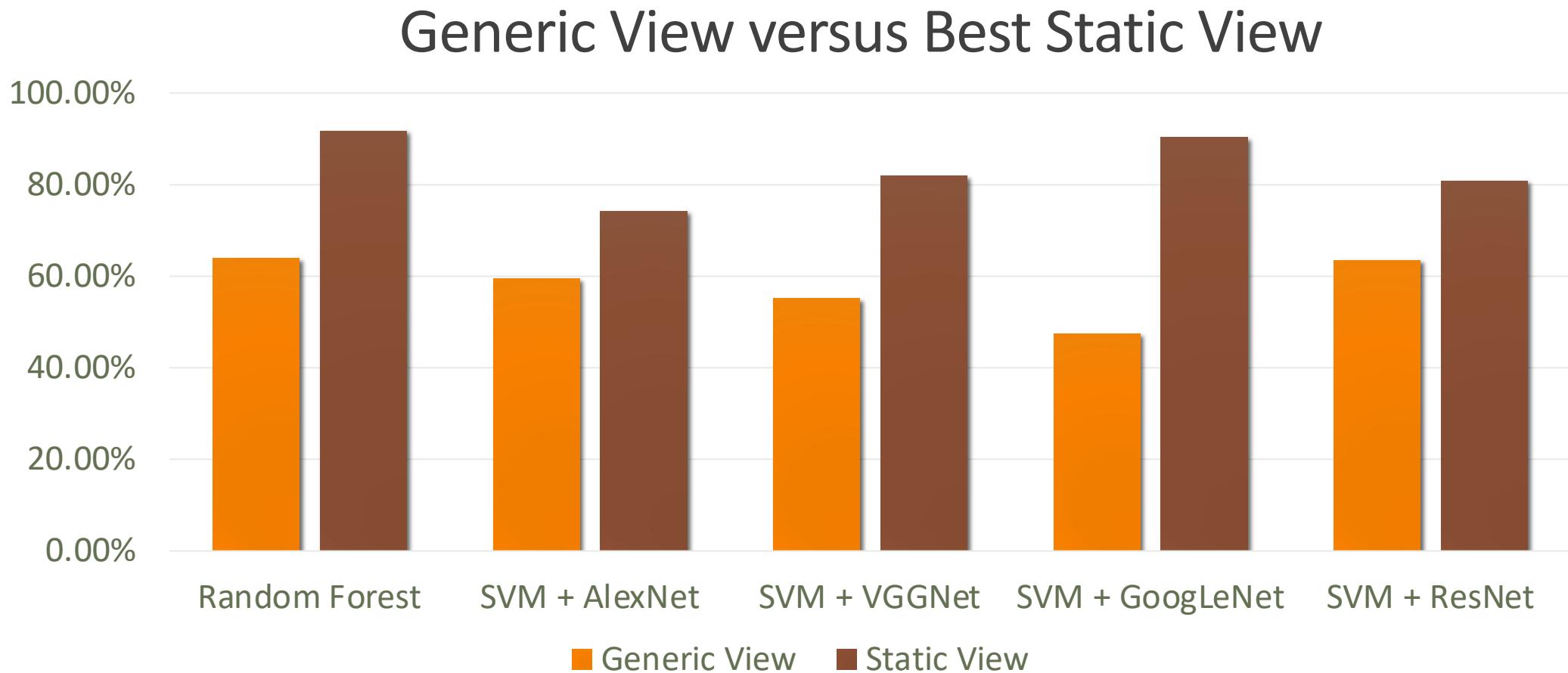


Ikea Table Assembly

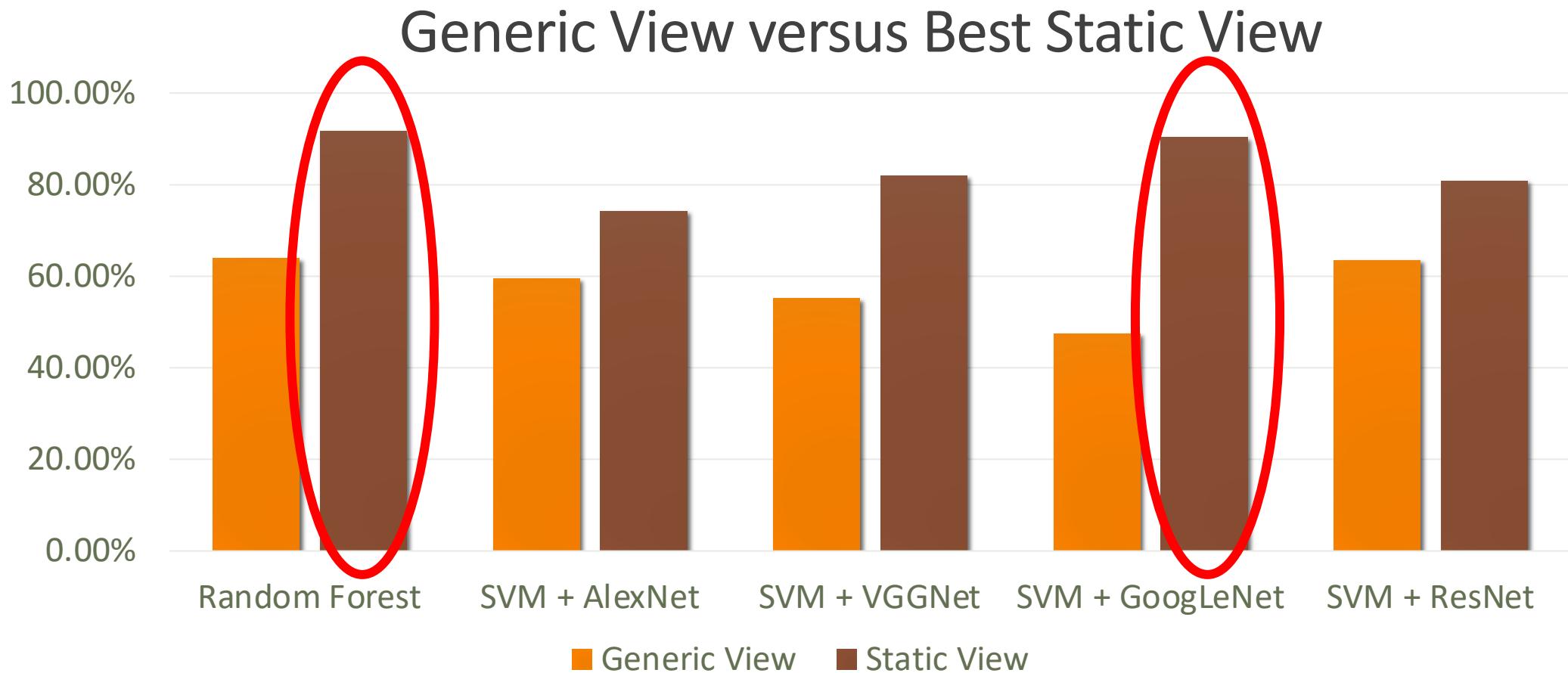
- PR2 attempts to insert a screw of a table leg into a predrilled hole on the table base
- 18 candidate viewpoints + “generic” viewpoint
- 39 executions of the task



Ikea Table Assembly Task

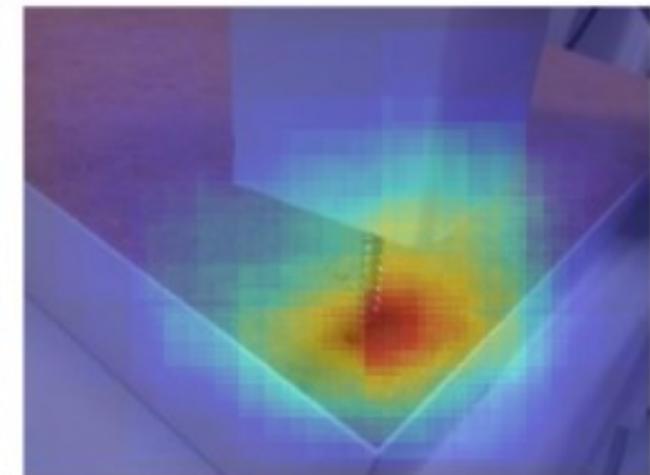
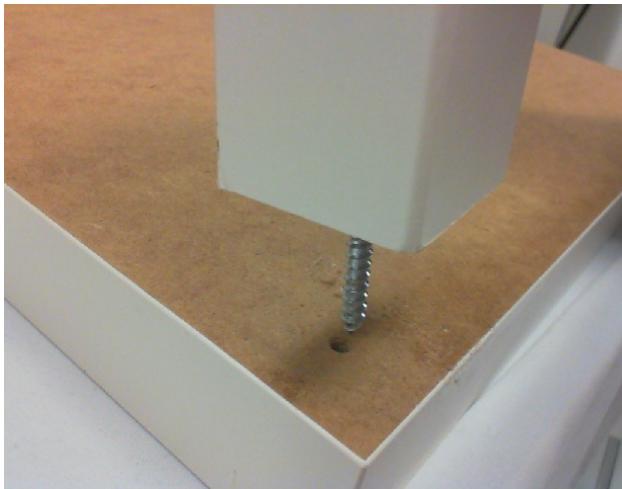


Ikea Table Assembly Task

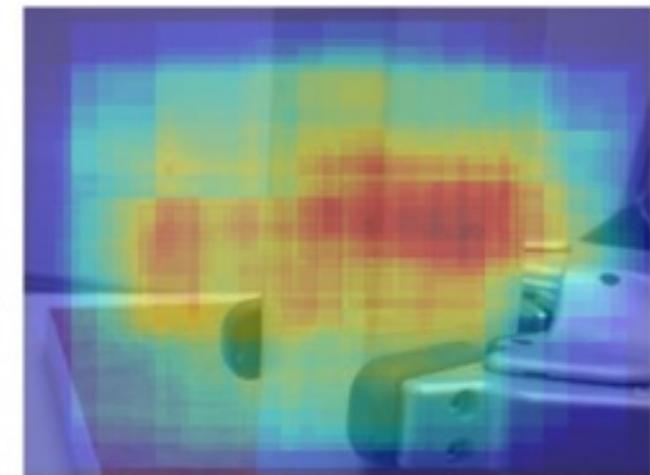


Ikea Table Assembly Task

Best Static View



Worst Static View



Experimental Domains

- 1. Ikea Table Assembly Task**
 - a. Without Obstacles
 - b. **With Dynamic Obstacles**
2. Quadrotor Surveillance Task with Dynamic Human Location

Ikea Table Assembly with Obstacles

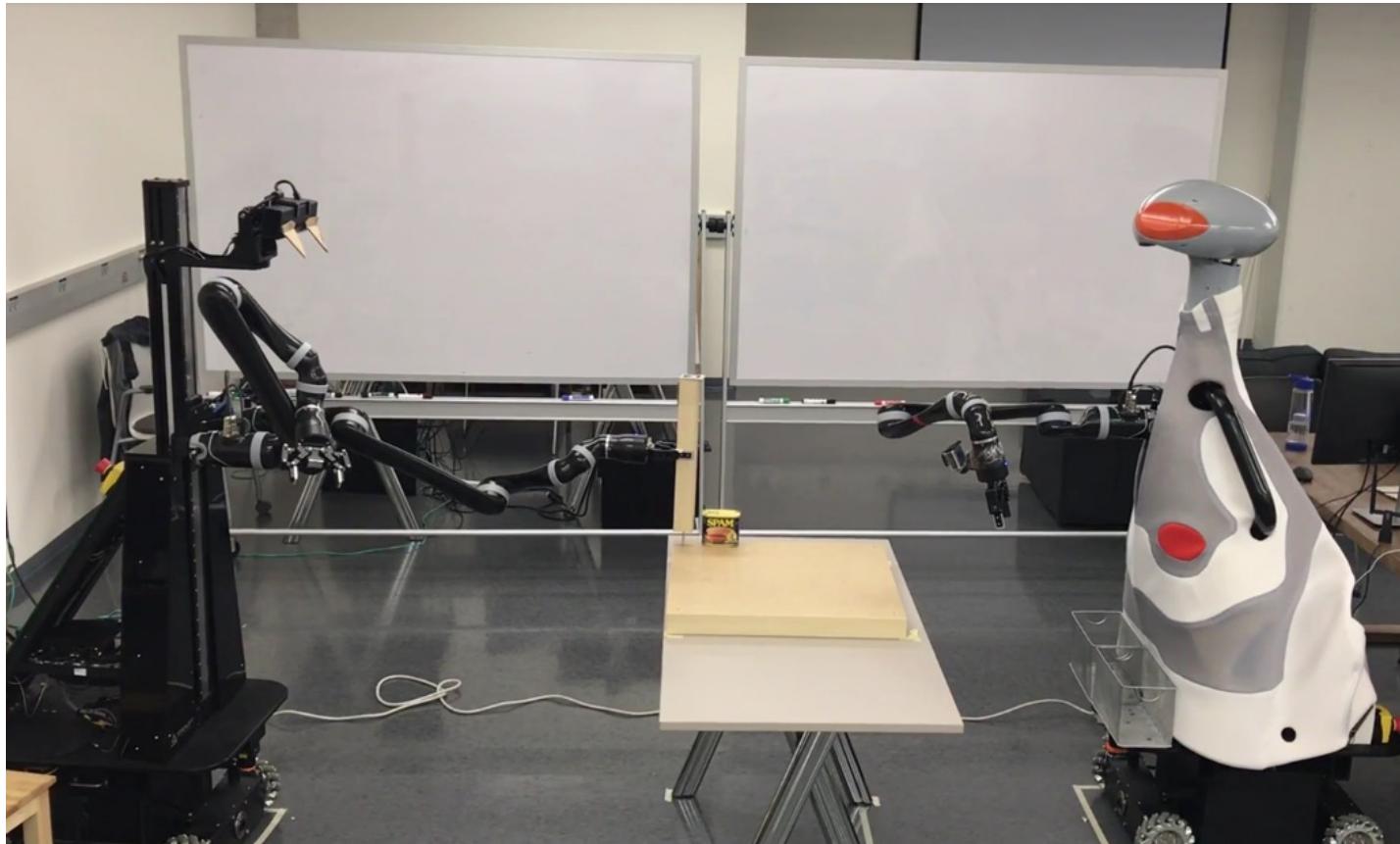
- 20 candidate viewpoints



- 40 executions of the task



Ikea Table Assembly with Obstacles



Ikea Table Assembly Task with Obstacles

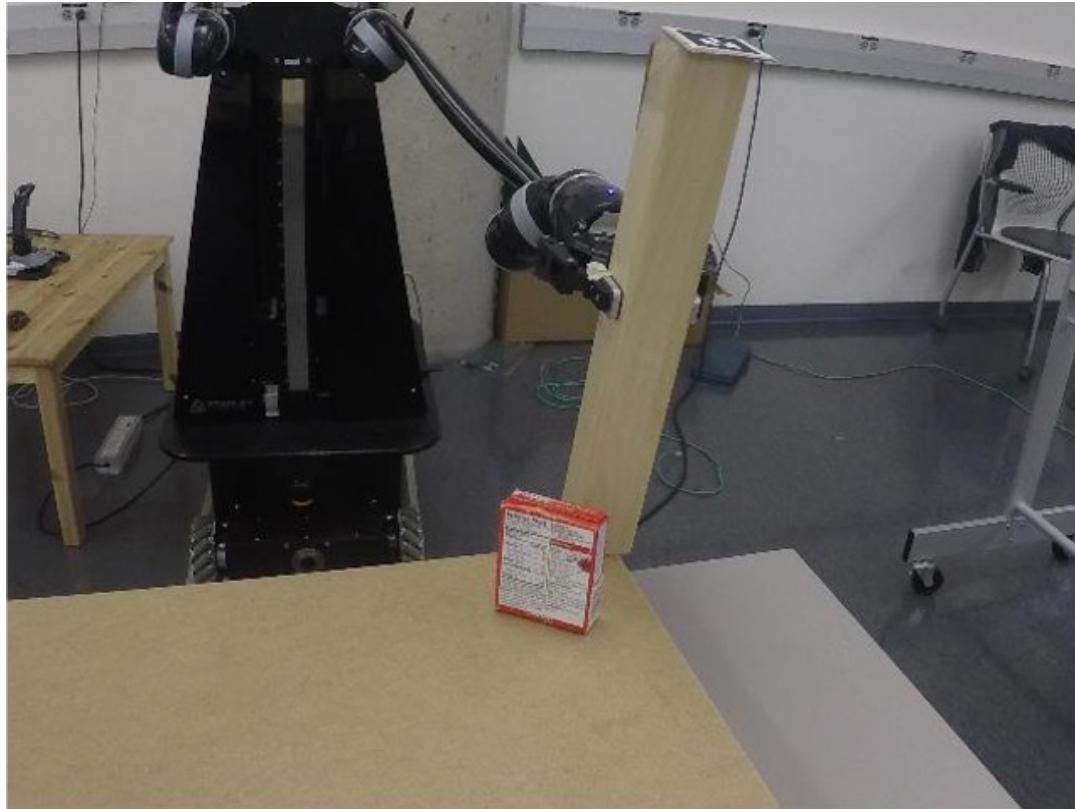


Best Static View



Active View

Ikea Table Assembly Task with Obstacles

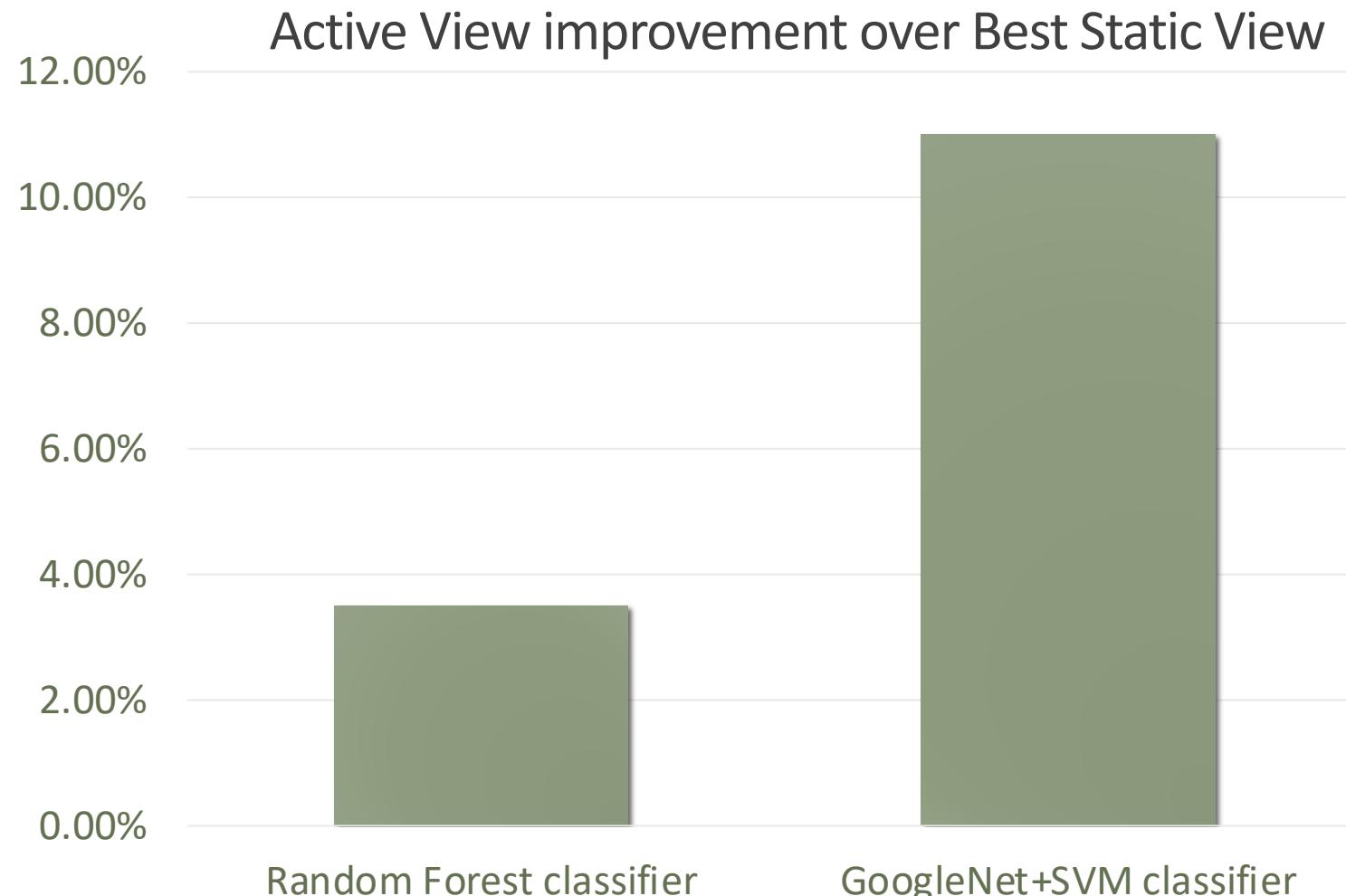


Best Static View



Active View

Ikea Table Assembly Task with Obstacles



Experimental Domains

1. Ikea Table Assembly Task
 - a. Without Obstacles
 - b. With Dynamic Obstacles
2. **Quadrotor Surveillance Task with Dynamic Human Location**

Quadrotor Surveillance

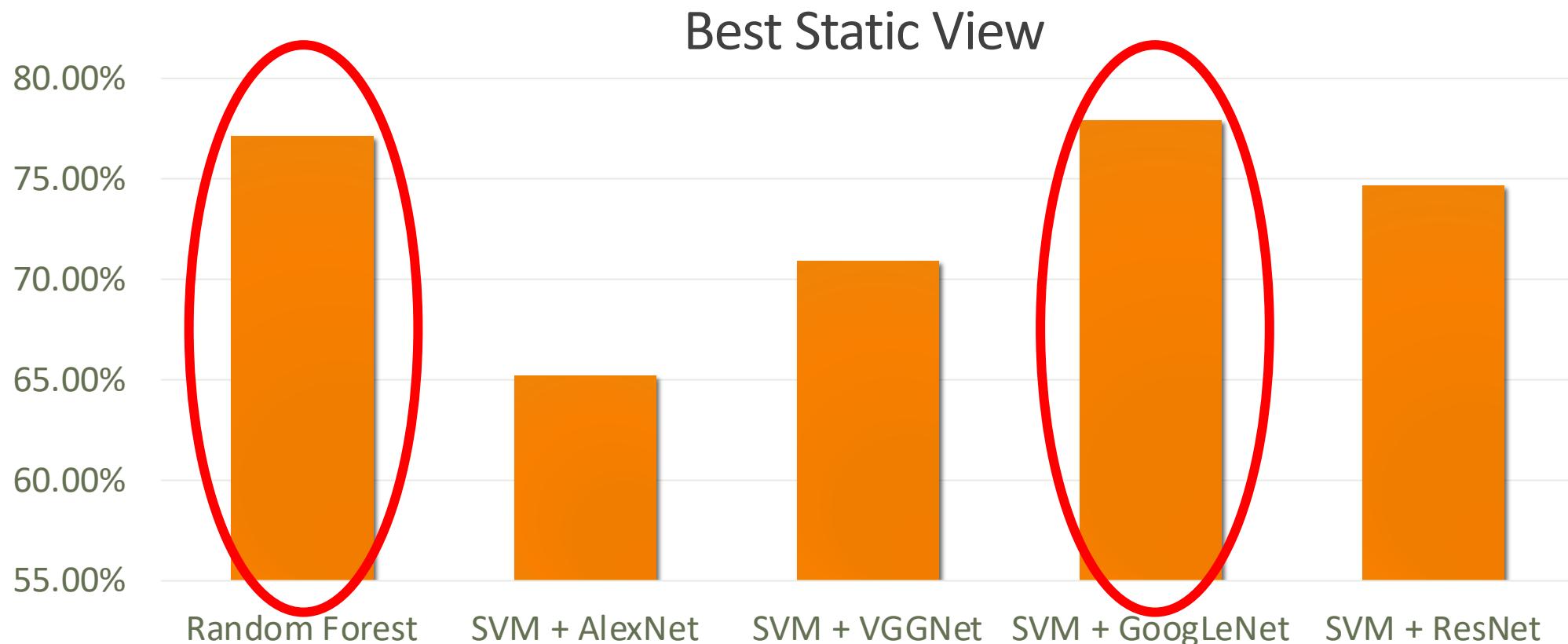
- Determine whether a person trying to “break-in” or standing next to the car



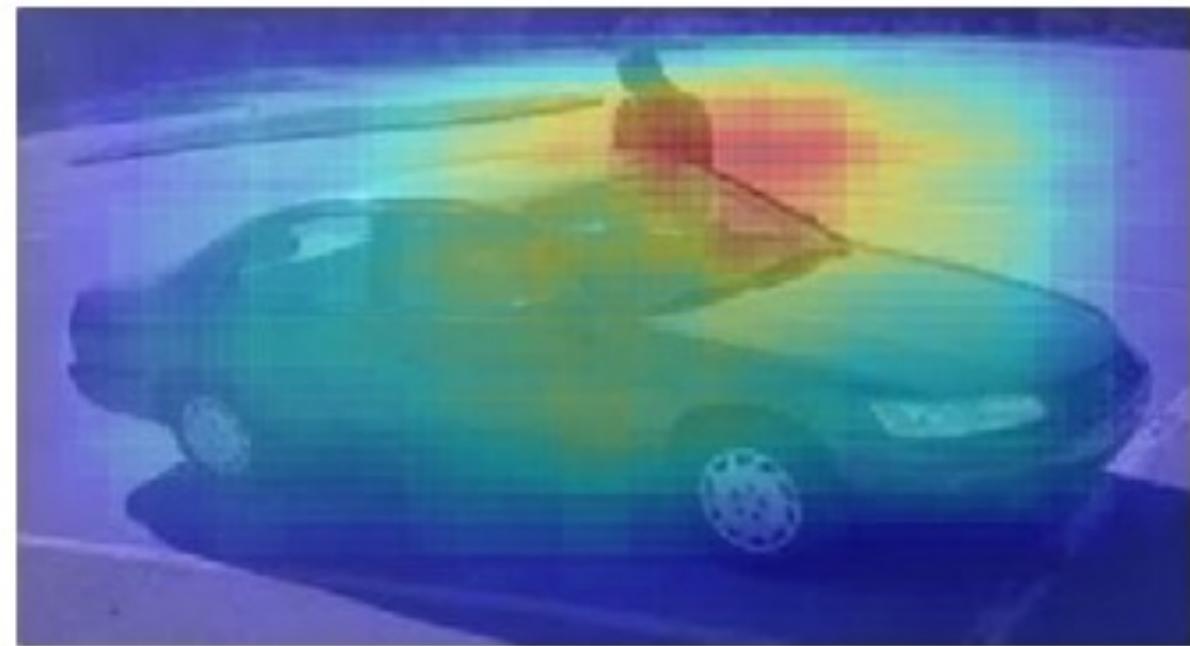
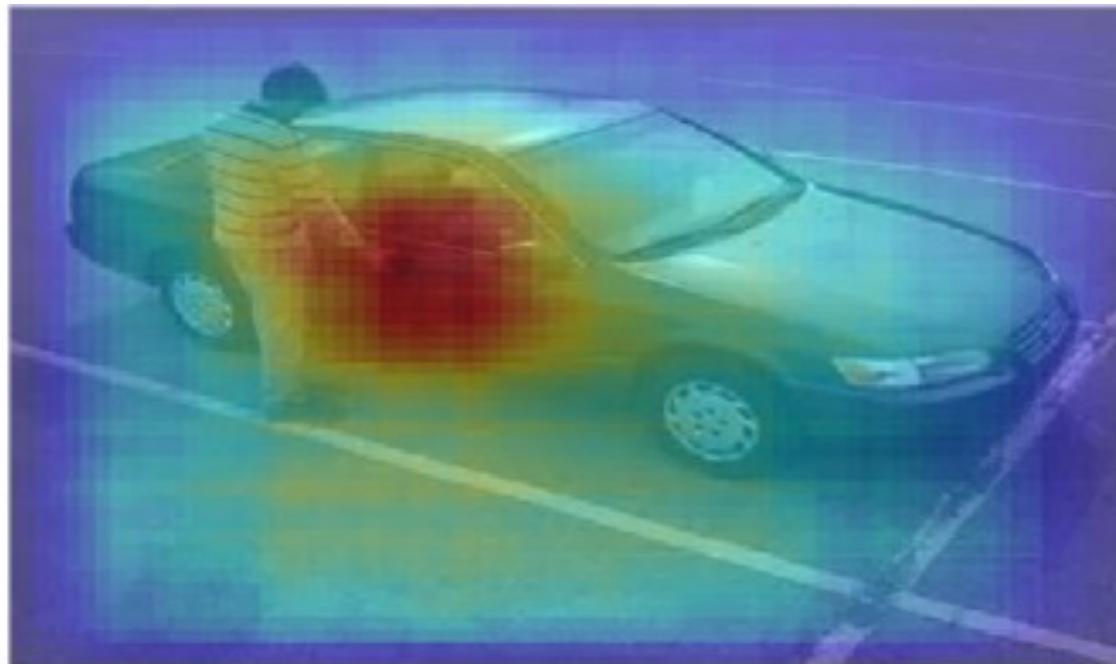
- 5 candidate viewpoints
- 68 executions of the task



Quadrotor Surveillance Task

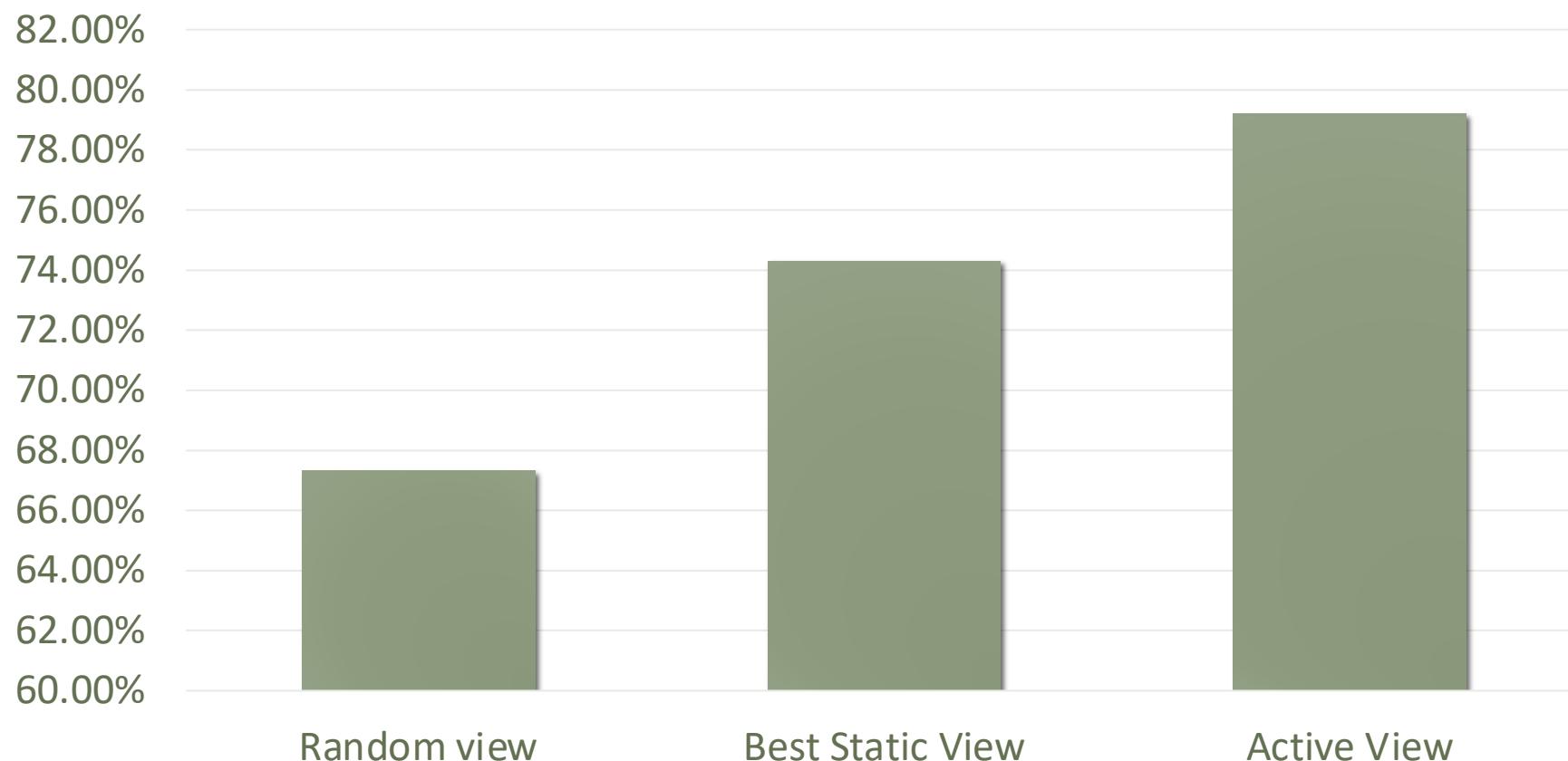


Quadrotor Surveillance Task



Quadrotor Surveillance Task

Accuracy of outcome classification



Conclusion

- Viewpoint selection is both possible and beneficial on several types of robotics tasks
- Potential of CNN features (extracted by using pre-trained networks) to perform surprisingly well on this problem, despite not being designed to extract fine-grained features
- Performance can be further improved by actively selecting a viewpoint based on the image content.

Thank You!

Questions?