AI4IA
AI for Industrial Assembly

Anton Agafonov - Nimrod Curtis - Sher Hazan

Our team at **Bosch Corporate Research** develops end-to-end solutions for robotic manipulation, leveraging state-of-the-art AI techniques to tackle industrial assembly challenges.

AI4IA enables precise, scalable, and robust robotic assembly, ensuring adaptability to real-world industrial environments.

Perception Utils

Object Tracker

- ✓ Real-time streaming at 25 FPS
- ✓ Zero-shot for novel objects
- ✓ Multi-object tracking
- ✓ Based on state-of-the-art model (SAM2)
- ✓ Robust to high-speed robot motion
- ✓ Initialized using prior data (prompt, frame)

6D Pose Estimator

- ✓ Real-time streaming at 10 FPS
- ✓ Zero-shot for novel objects
- ✓ Switching between objects and views
- ✓ Based on state-of-the-art models (SAM2, MegaPose)
- ✓ Robust for high-speed robot motion
- ✓ Initialized by prior data (prompt, pose, frame)

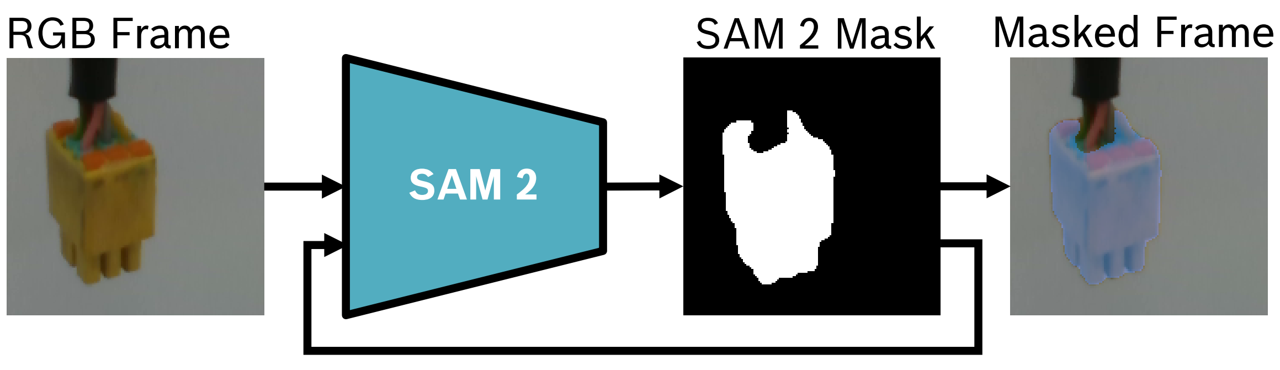
Behavior Cloning

- ✓ Deterministic Policy (DAgger variation)
- ✓ Generalizes across diverse grasping scenarios
- ✓ Robust to varying lighting conditions
- ✓ Handles initial plug pose variations efficiently
- ✓ Real-time inference (ResNet-18 architecture)

Object Detector

- ✓ Simple and fast onboarding of novel objects
- ✓ Zero-shot detection for novel objects
- ✓ Built on state-of-the-art models (SAM2, Dinov2)
- ✓ Robust to occlusions and varying lighting conditions
- ✓ Estimates object coarse pose
- ✓ CLS tokens as descriptors for object representation

Object Tracker



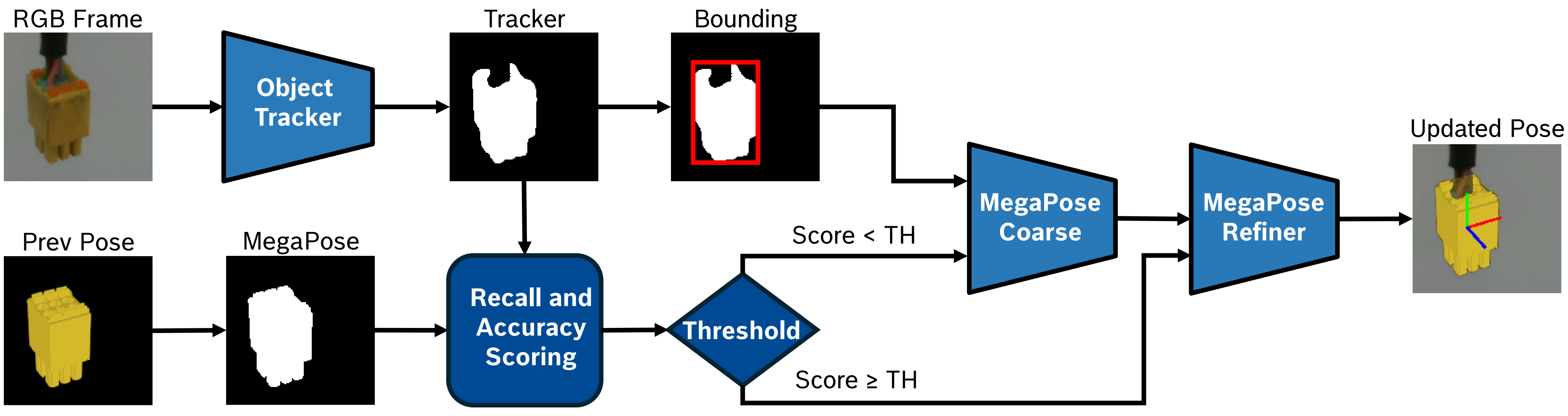
Behavior Cloning

Data collecting and training algorithm:

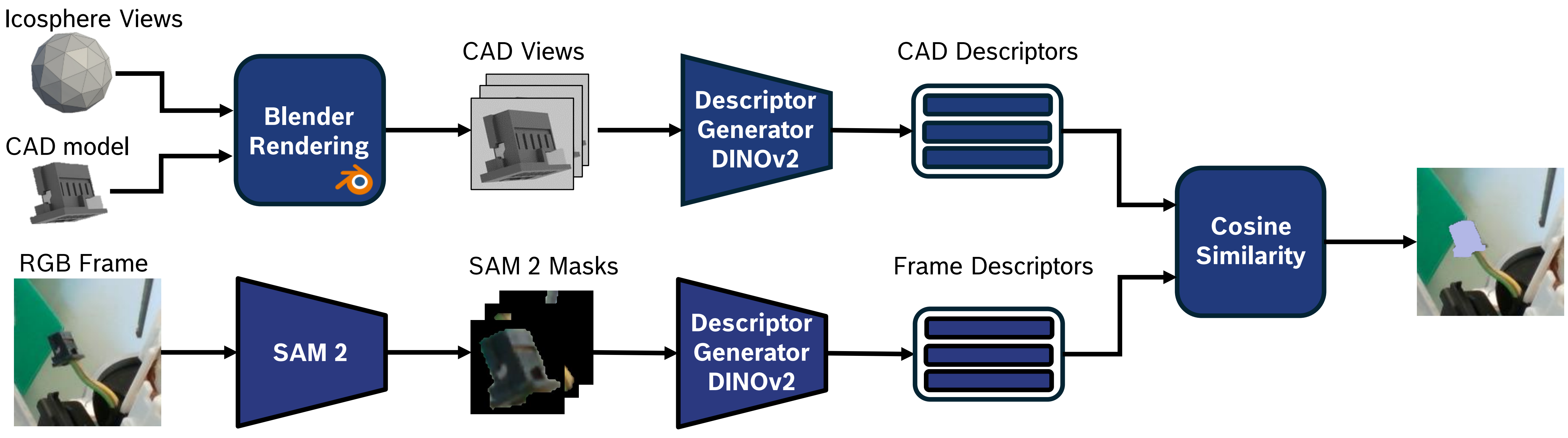
```
Initialize $\mathcal{D} \leftarrow \emptyset$
Initialize expert policy $\boldsymbol{\pi}_{expert}$ from vector field
Initialize agent policy $\boldsymbol{\pi}_{agent\_0}$ as NN with Resnet18 backbone
Initialize $\gamma$

for $i = 0$ to $N$ do
    $\beta_t = \max(1 - \gamma * t, 0)$
    for $j = 1$ to $T$ do
        $\boldsymbol{\pi}(\boldsymbol{S}) = \begin{cases} \boldsymbol{\pi}_{expert}(\boldsymbol{s})\ \boldsymbol{w.p.}\ \beta_t \\ \boldsymbol{\pi}_{agent\_i}(\boldsymbol{s})\ \boldsymbol{w.p.}\,(1-\beta_t) \end{cases}$
        Perform action $\boldsymbol{A} = \boldsymbol{\pi}(\boldsymbol{s})$
        Add to $\mathcal{D}_i = \{(s, \boldsymbol{\pi}_{expert}(\boldsymbol{s}))\}$
    end for
    Aggregate datasets: $\mathcal{D} \leftarrow \mathcal{D} \cup \mathcal{D}_i$
    Train agent policy $\boldsymbol{\pi}_{agent\_i}$ on $\mathcal{D}$
end for
```

6D Pose Estimator



Scene Object Detector



Robotic Skills

Simple Move

- ✓ Basic skill for moving the robot to a predefined destination
- ✓ Admittance control adjusts motion based on external forces
- ✓ Minimum jerk trajectory planning for natural motion

Move to Object

- ✓ Utilizes 6D Pose Estimator
- ✓ Estimates object`s pose based on EEF motion
- ✓ Filters object poses in the base frame
- ✓ Moves EEF to align with the object`s estimated location

Grasp

- ✓ Closes the gripper with force-based control
- ✓ Ensures secure grasp
- ✓ Prevents object damage

Screw / Unscrew

- ✓ Screw / Unscrew bolts with force-based control
- ✓ PID force control maintains constant bolt pressure
- ✓ Force feedback to detect success or failure

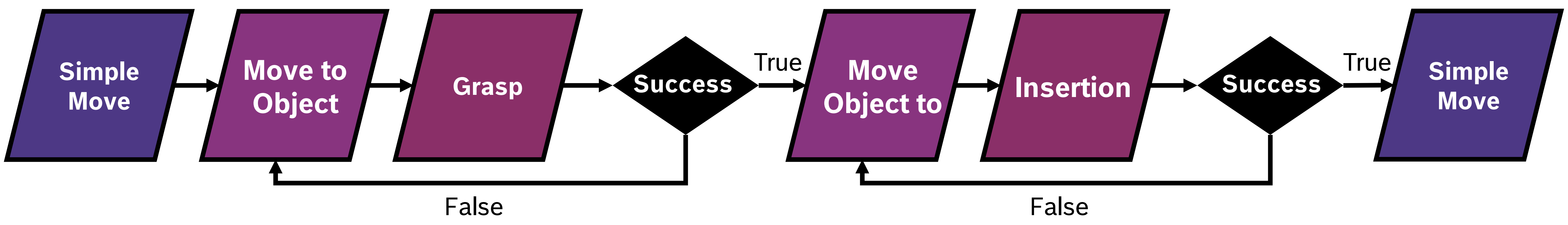
Move Object to

- ✓ Utilizes 6D Pose Estimator
- ✓ Filters object poses in EEF frame
- ✓ Moves the object to a defined pose in the base frame

Insertion

- ✓ Insertion policy controls for smooth insertion
- ✓ Spiral search reset in case of insertion failure
- ✓ Behavior cloning-based refinement

Behavior Tree



Future Work

Perception Utils

- ❑ Fuse Depth modality in the Perception Utils
- ❑ Enhance multi-view perception
- ❑ Optimize real-time inference to achieve higher FPS
- ❑ Research and deployment of new state-of-the-art models

Robotic Skills

- ❑ Adaptive Behavior Tree generation utilizing VLA
- ❑ Online refinement of insertion policy
- ❑ Learn manipulation policies in simulation
- ❑ Deploy Sim2Real policies

REFERENCES

1. S. Ross, G. Gordon, and J. A. Bagnell, “A reduction of imitation learning and structured prediction to no-regret online learning.
2. Labbé, Y., Manuelli, L., Mousavian, A., Tyree, S., Birchfield, S., Tremblay, J., Carpentier, J., Aubry, M., Fox, D., & Sivic, J. (2022). *MegaPose: 6D Pose Estimation of Novel Objects via Render & Compare.*
3. Ravi, N., Gabeur, V., Hu, Y.-T., Hu, R., Ryali, C., Ma, T., Khedr, H., Rädle, R., Rolland, C., Gustafson, L., Mintun, E., Pan, J., Alwala, K. V., Carion, N., Wu, C.-Y., Girshick, R., Dollár, P., & Feichtenhofer, C. (2024). SAM 2: Segment Anything in Images and Videos.
4. V. N. Nguyen, T. Groueix, G. Ponimatkin, V. Lepetit, and T. Hodan, "CNOS: A Strong Baseline for CAD-based Novel Object Segmentation