ECCV'20 ONLINE

23-28 AUGUST 2020

16TH EUROPEAN CONFERENCE ON

COMPUTER VISION

WWW.ECCV2020.EU

ECCV'20 ONLINE
23-28 AUGUST 2020

How to track your dragon:

A Multi-Attentional Framework for real-time RGB-D Object Pose Tracking

Isidoros Marougkas

Petros Koutras

Nikos Kardaris

Georgia Chalvatzaki

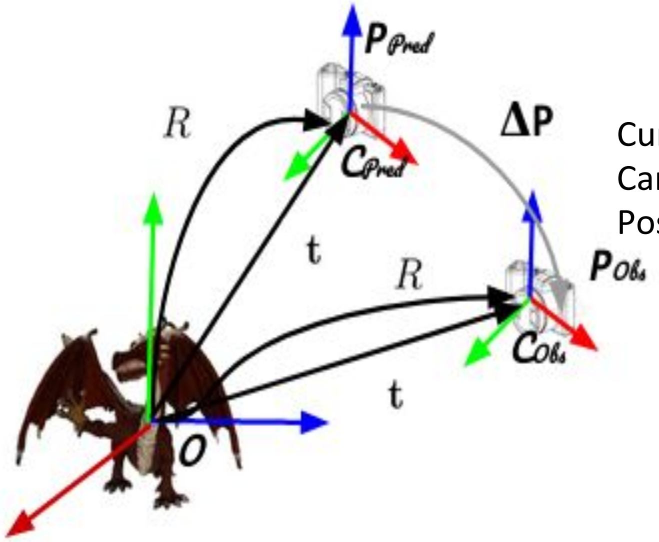
Georgios Retsinas

Petros Maragos

TECHNISCHE UNIVERSITÄT DARMSTADT

Problem Definition/Pose Space:

Previous Camera Pose:



Current
Camera
Pose:

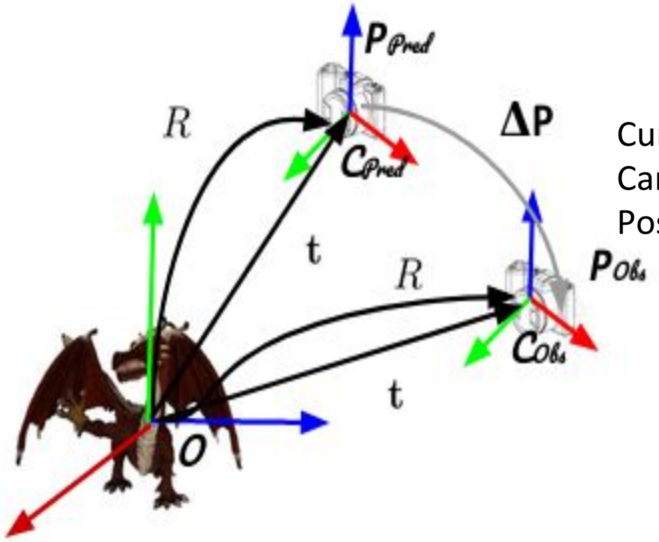
Problem Definition/Pose Space:

Classical definition of the Pose Space:

$$\mathscr{C} = \left\{ \mathbb{P} \;\middle|\; \mathbb{P} = \left[\begin{array}{c|c} R & \mathbf{t} \\ \hline \mathbf{0}^T & 1 \end{array}\right], \mathbf{t} \in \mathbb{R}^3, R \in \mathrm{SO}(3) \right\}$$

neglects the objects' symmetries.

Previous Camera Pose:



Current Camera Pose:

Brégier et al.[2]: augmented it to account for this discrepancy:

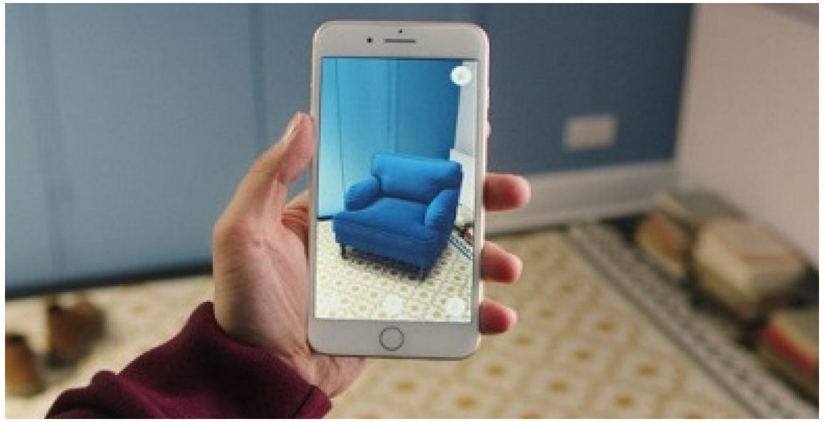
$$\mathscr{C} = \left\{ \mathbb{P} \;\middle|\; \mathbb{P} = \left[\begin{array}{c|c} R \cdot G & \mathbf{t} \\ \hline \mathbf{0}^T & 1 \end{array}\right], \mathbf{t} \in \mathbb{R}^3, R \in \mathrm{SO}(3), G \in \mathrm{SO}(3) \right\}.$$

<u>For asymmetrical objects:</u> $G = \mathbb{I}_3$

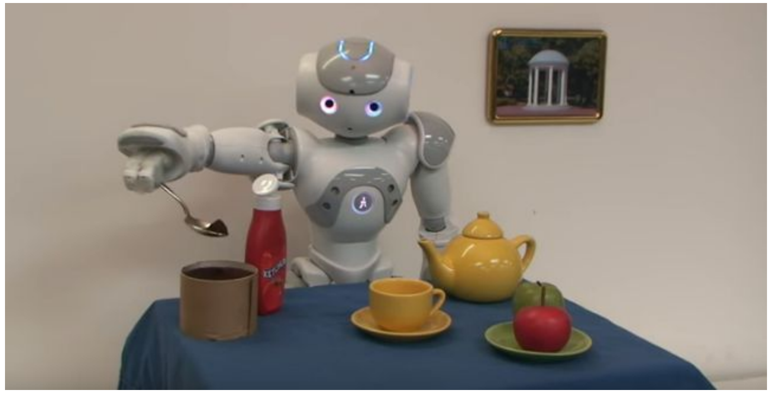
Brégier, R., Devernay, F., Leyrit, L., Crowley, J.L.: Defining the pose of any 3drigid object and an associated distance. Int. J. of Comp. Vision (IJCV)126(6),571–596 (2018)

Motivation: Applications

Augmented Reality:



Robotic Grasping & Manipulation:

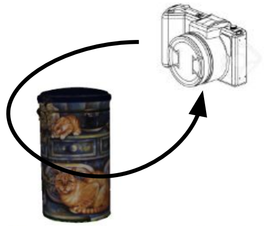


Autonomous Driving:

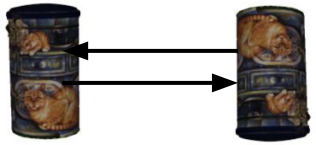
ECCV'20 ONLINE 23-28 AUGUST 2020

Challenges:

- ➢ Appearance change due to pose variation
- ➢ Modelling of sensor noise
- ➢ Illumination conditions
- ➢ Pose Ambiguities
  - ○ Rotation Representation
  - ○ Object Symmetries
    - ■ Continuous (Rotational)



    - ■ Discrete (Reflective)



- ➢ Motion blur
- ➢ Object size & texture
- ➢ Background Clutter - Color Noise
- ➢ Occlusions
  - ○ Static

  - ○ Dynamic

- ➢ Pose drift accumulated over time

ECCV'20 ONLINE 23-28 AUGUST 2020

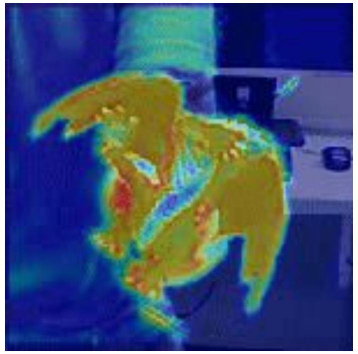
‘Hard Interaction’ scenario:

<u>Qualitative Results:</u> re-iterate every time the tracker fails irrecoverably

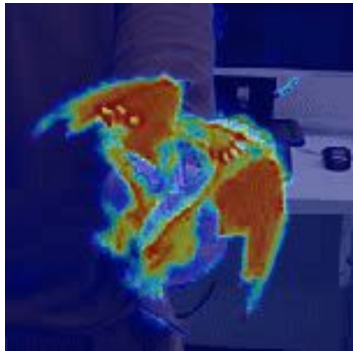
Garon et al.[8]

Ours

Foreground Attention:



Occlusion Attention:



Garon, M., Laurendeau, D., Lalonde, J.F.: A framework for evaluating 6-dof objecttrackers. In: Proc. European Conf. on Computer Vision (ECCV). pp. 582–597(2018)

Contributions:

- **Spatial Attention** mechanism for **Background Clutter and Occlusion Handling**
    - Supervision: extracted by fully exploiting the synthetic nature of our training data
    - Provides intuitive understanding of the tracker's region of interest

Contributions:

- **Spatial Attention** mechanism for **Background Clutter and Occlusion Handling**
  - Supervision: extracted by fully exploiting the synthetic nature of our training data
  - Provides intuitive understanding of the tracker's region of interest

- **Multi-Task Pose Tracking Loss** function that:
  - Respects the geometry
    - of the Object's 3D model

      and
    - of the Pose Space

Contributions:

- **Spatial Attention** mechanism for **Background Clutter and Occlusion Handling**
  - Supervision: extracted by fully exploiting the synthetic nature of our training data
  - Provides intuitive understanding of the tracker's region of interest

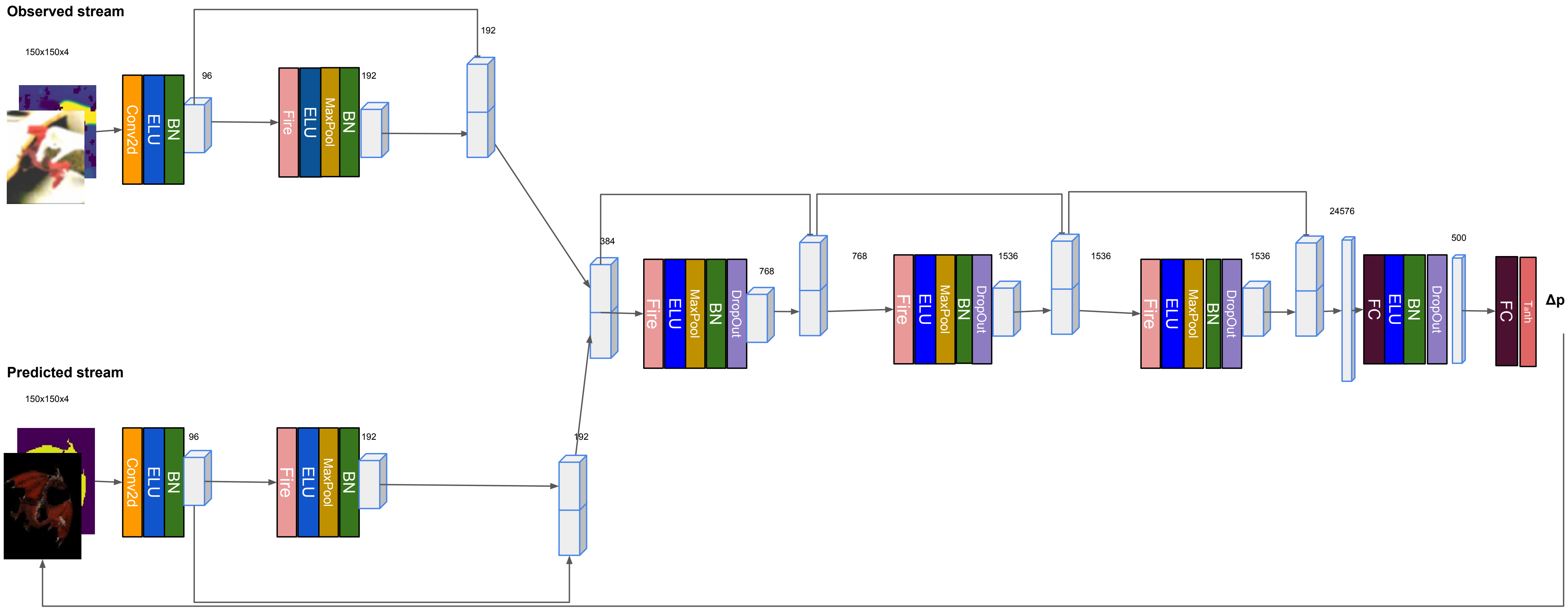
- **Multi-Task Pose Tracking Loss** function that:
  - Respects the geometry
    - of the Object's 3D model

      and
    - of the Pose Space

- **SoA real-time** performance in the hardest scenarios of Garon et al.[8]
  - **34.03%** drop in <u>Translation error</u> & **40.01%** drop in <u>Rotation error</u>

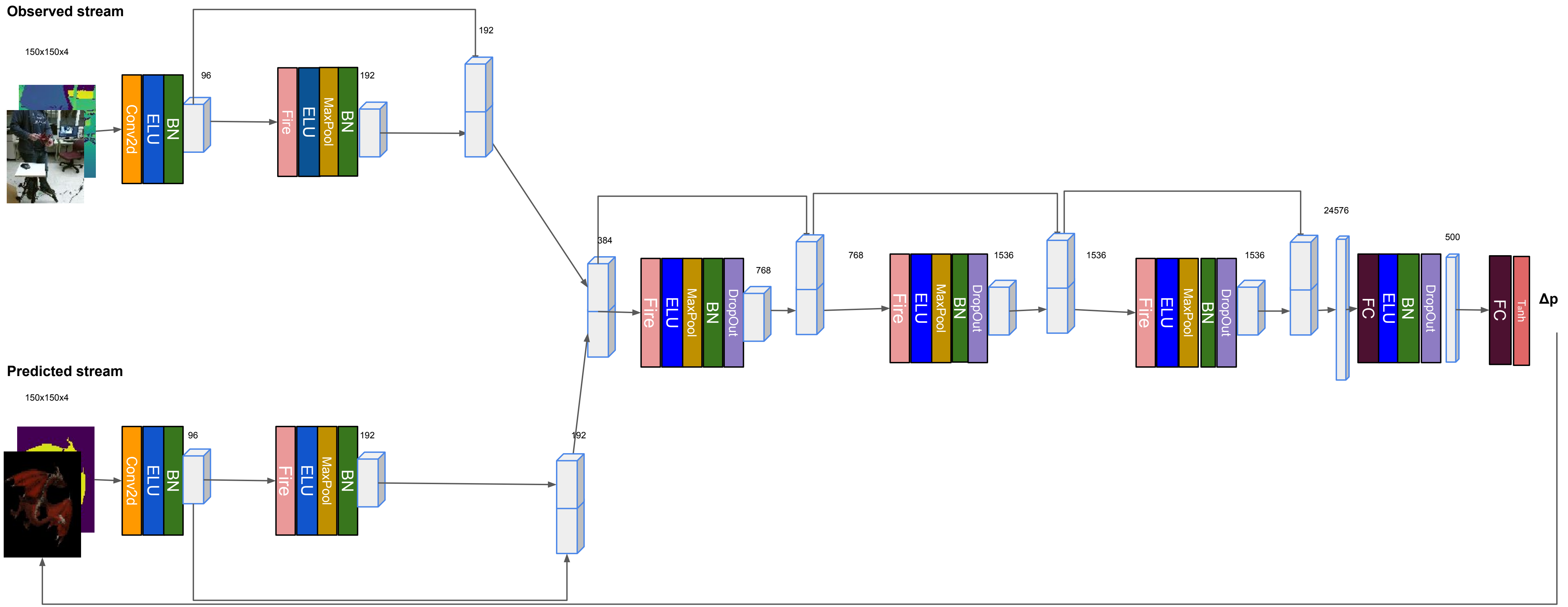
Garon, M., Laurendeau, D., Lalonde, J.F.: A framework for evaluating 6-dof objecttrackers. In: Proc. European Conf. on Computer Vision (ECCV). pp. 582–597(2018)

Baseline Architecture of Garon et al.[8] *(training mode)*:



Garon, M., Laurendeau, D., Lalonde, J.F.: A framework for evaluating 6-dof objecttrackers. In: Proc. European Conf. on Computer Vision (ECCV), pp. 582–597(2018)

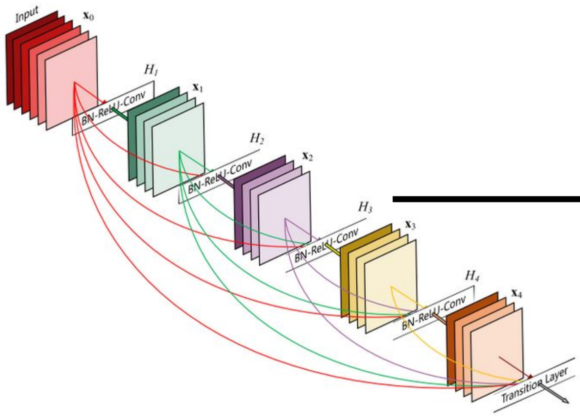
Baseline Architecture of Garon et al.[8] *(inference mode)*:

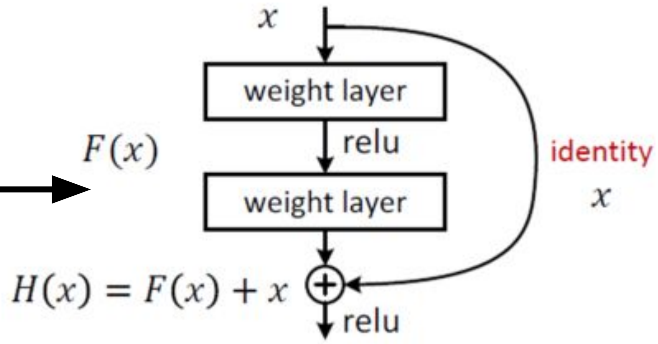


Garon, M., Laurendeau, D., Lalonde, J.F.: A framework for evaluating 6-dof objecttrackers. In: Proc. European Conf. on Computer Vision (ECCV). pp. 582–597(2018)

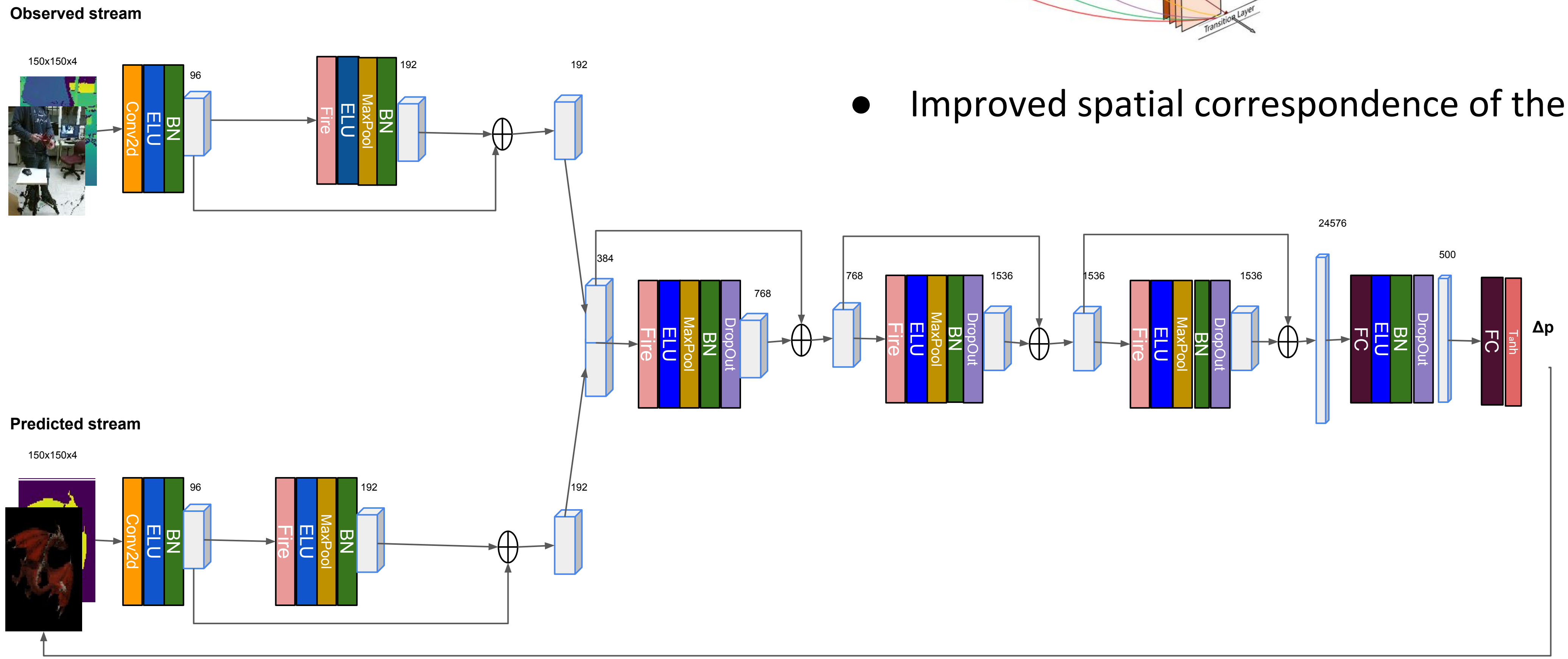
Our Architecture:

'Dense' ⟶ Residual connections:



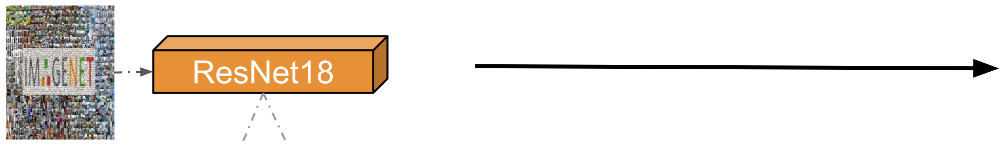


- Improved spatial correspondence of the features

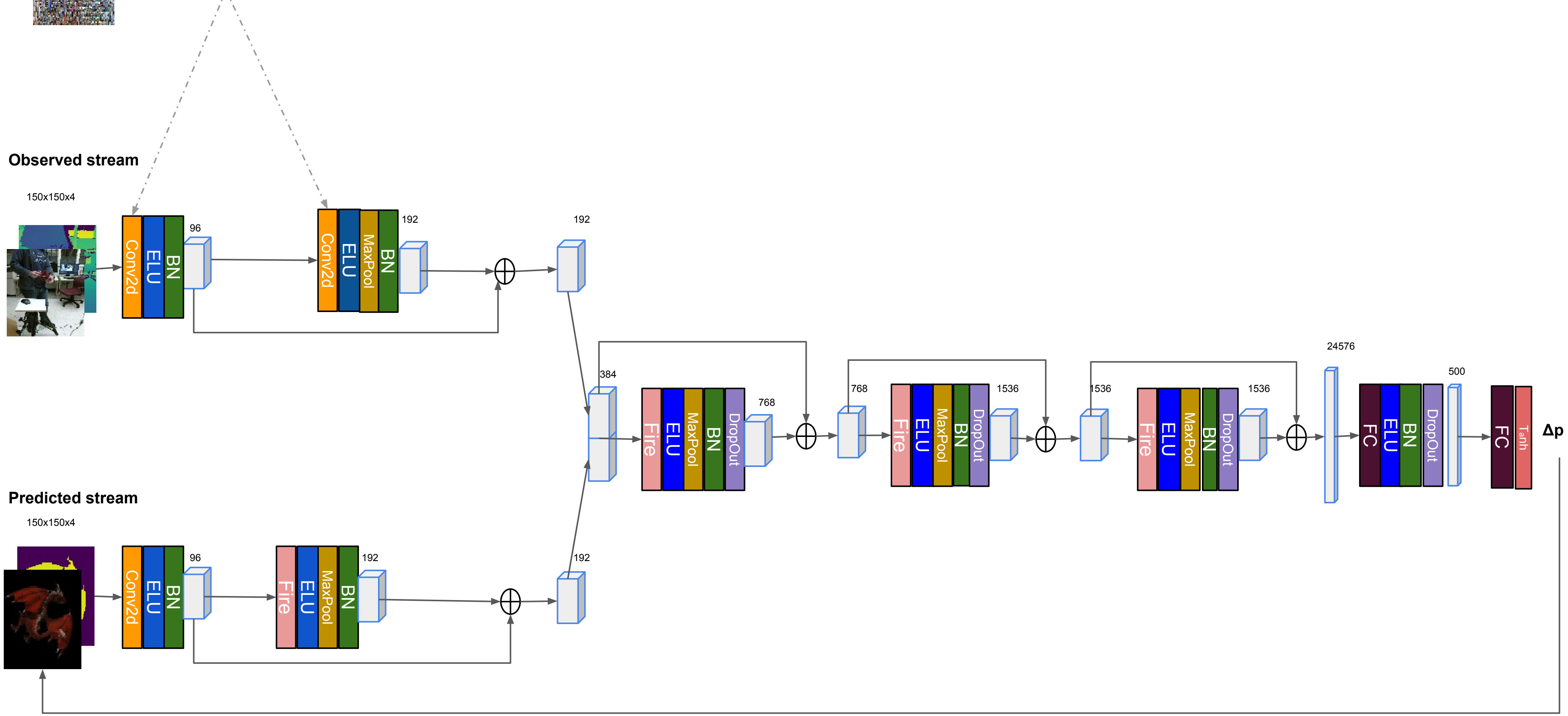


He, K., Zhang, X., Ren, S., Sun, J.: Deep residual learning for image recognition.In: Proc. IEEE Conf. on Computer Vision and Pattern Recognition (CVPR). pp.770–778 (2016)

Our Architecture:



- <u>Hintentoisser et al.[14]:</u> Bridging domain gap

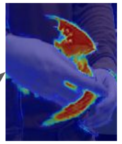


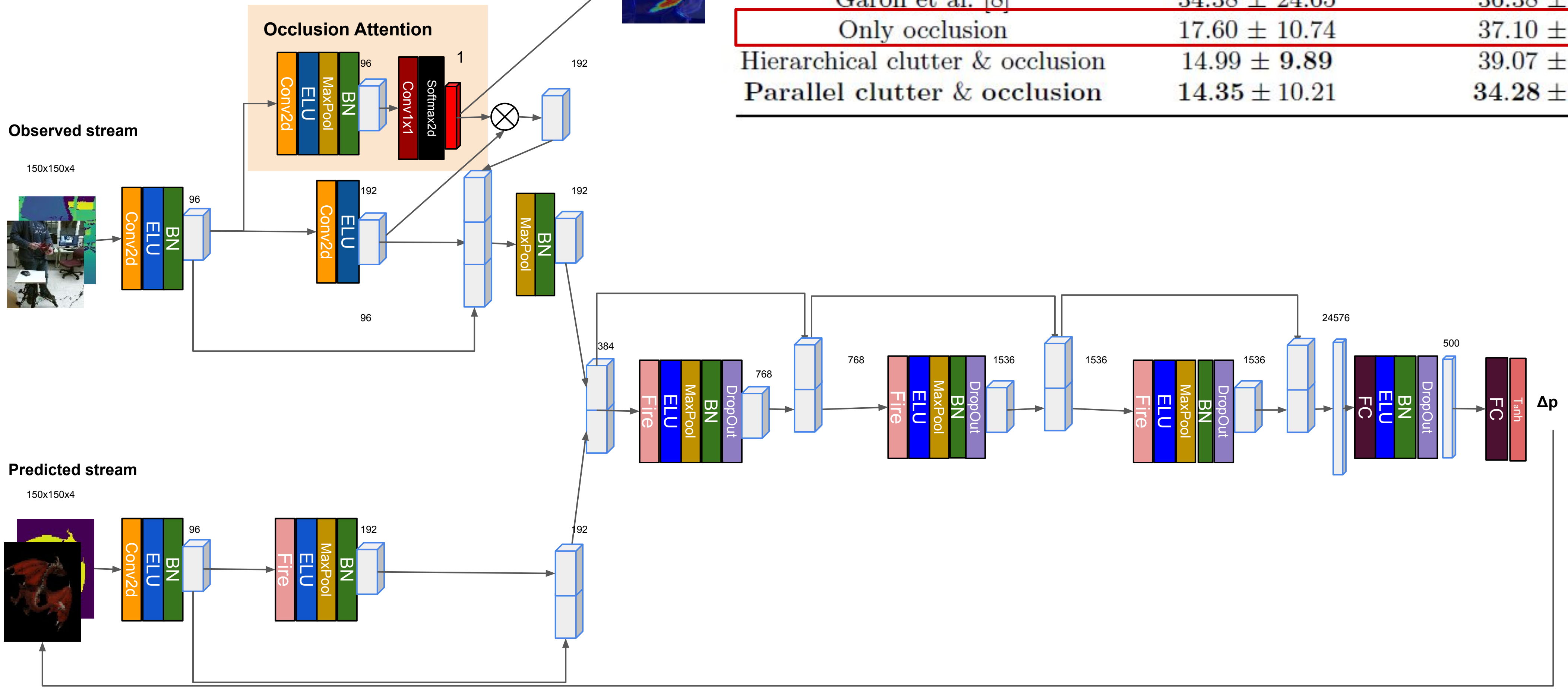
Hinterstoisser, S., Lepetit, V., Wohlhart, P., Konolige, K.: On pre-trained imagefeatures and synthetic images for deep learning. In: Proc. European Conf. on Com-puter Vision (ECCV). pp. 0–0 (2018)

Spatial Attention maps

(Only Occlusion Handling):

	Translational Error(mm)	Rotational Error(degrees)
Garon et al. [8]	34.38 ± 24.65	36.38 ± 36.31
Only occlusion	17.60 ± 10.74	37.10 ± 35.08
Hierarchical clutter & occlusion	14.99 ± 9.89	39.07 ± 33.22
Parallel clutter & occlusion	**14.35** ± 10.21	**34.28 ± 29.81**

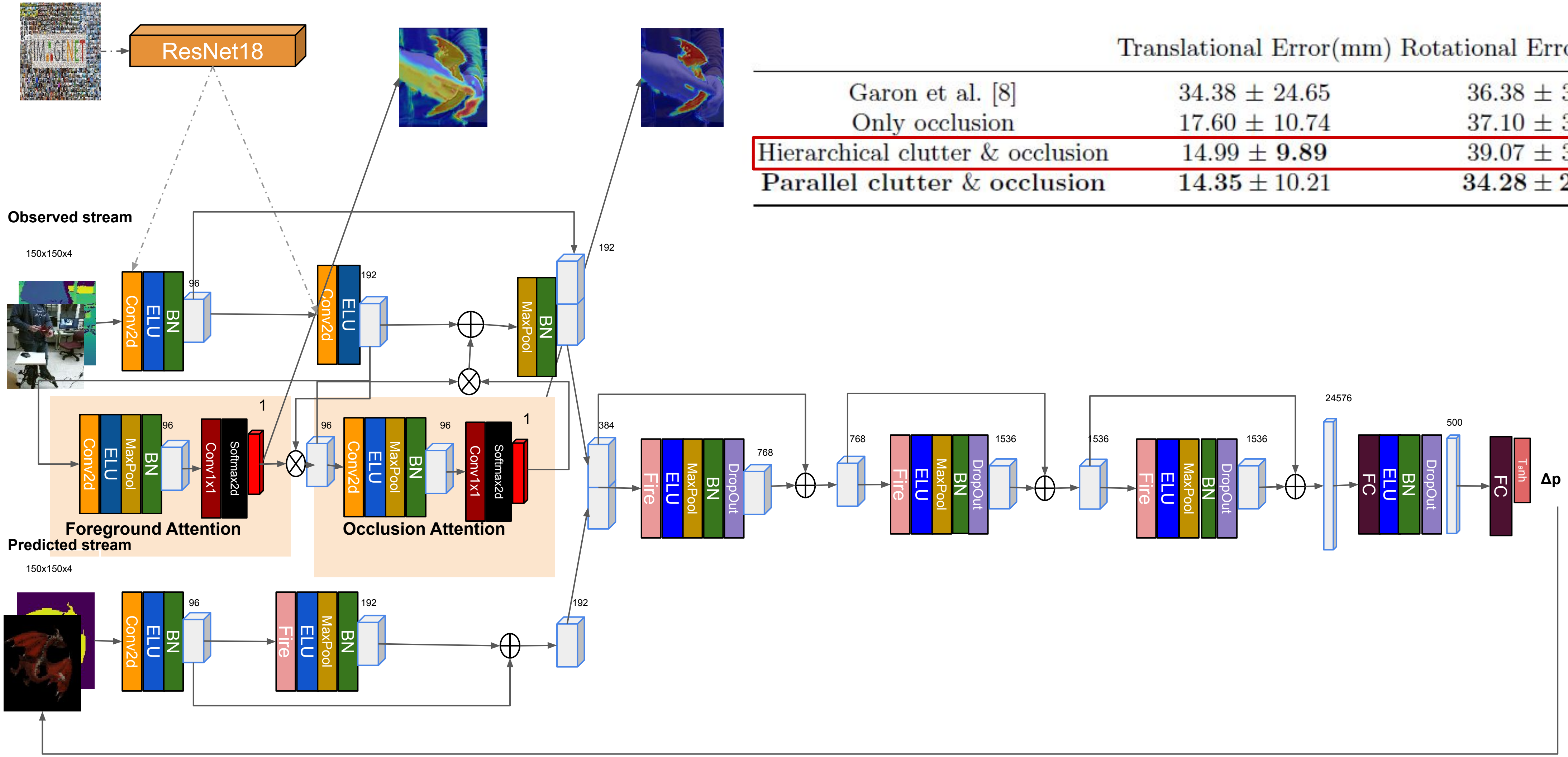




Spatial Attention maps

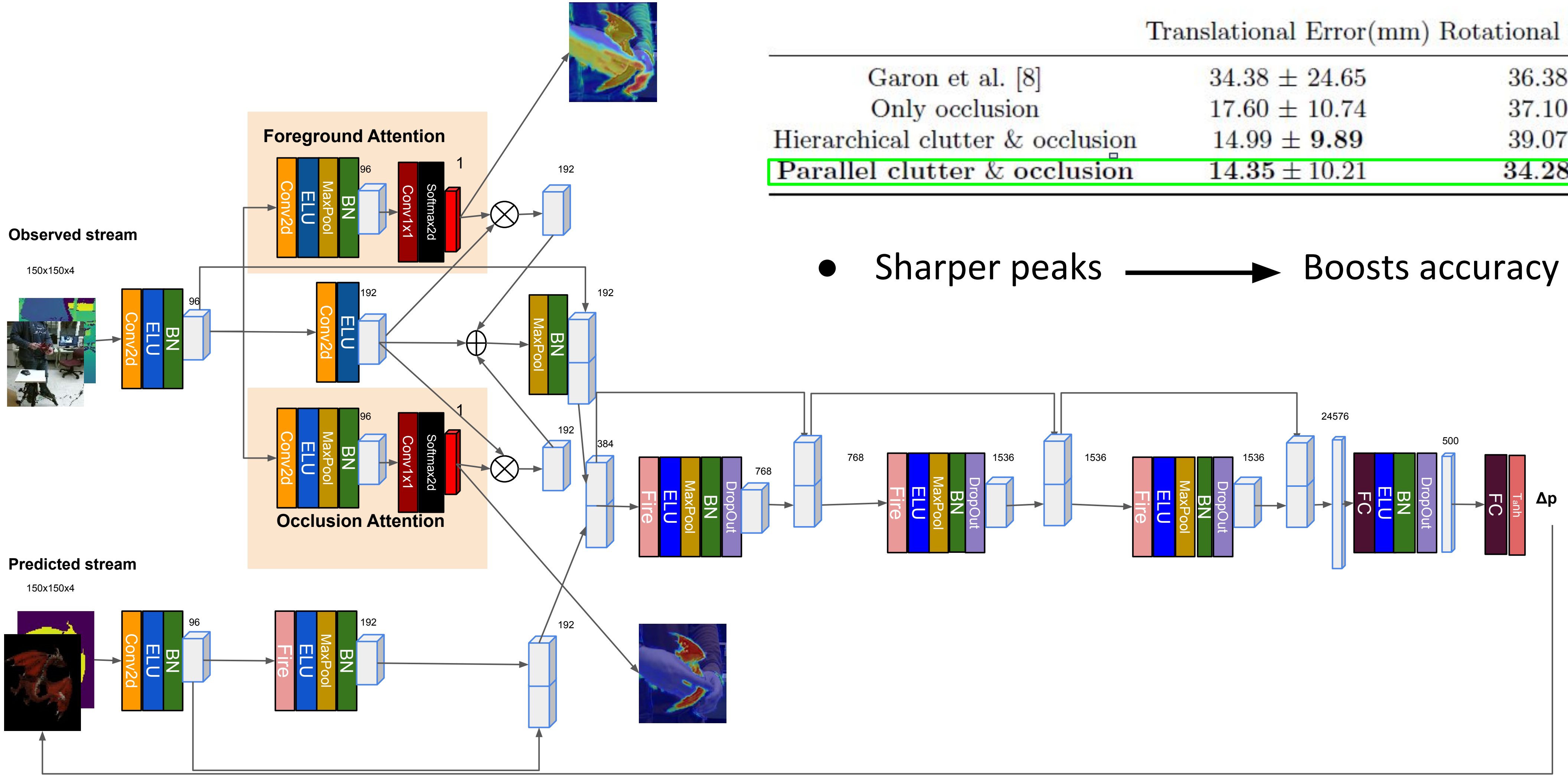
(Hierarchical connection):

	Translational Error(mm)	Rotational Error(degrees)
Garon et al. [8]	34.38 ± 24.65	36.38 ± 36.31
Only occlusion	17.60 ± 10.74	37.10 ± 35.08
Hierarchical clutter & occlusion	14.99 ± 9.89	39.07 ± 33.22
Parallel clutter & occlusion	**14.35 ± 10.21**	**34.28 ± 29.81**



Spatial Attention maps

(Parallel connection):

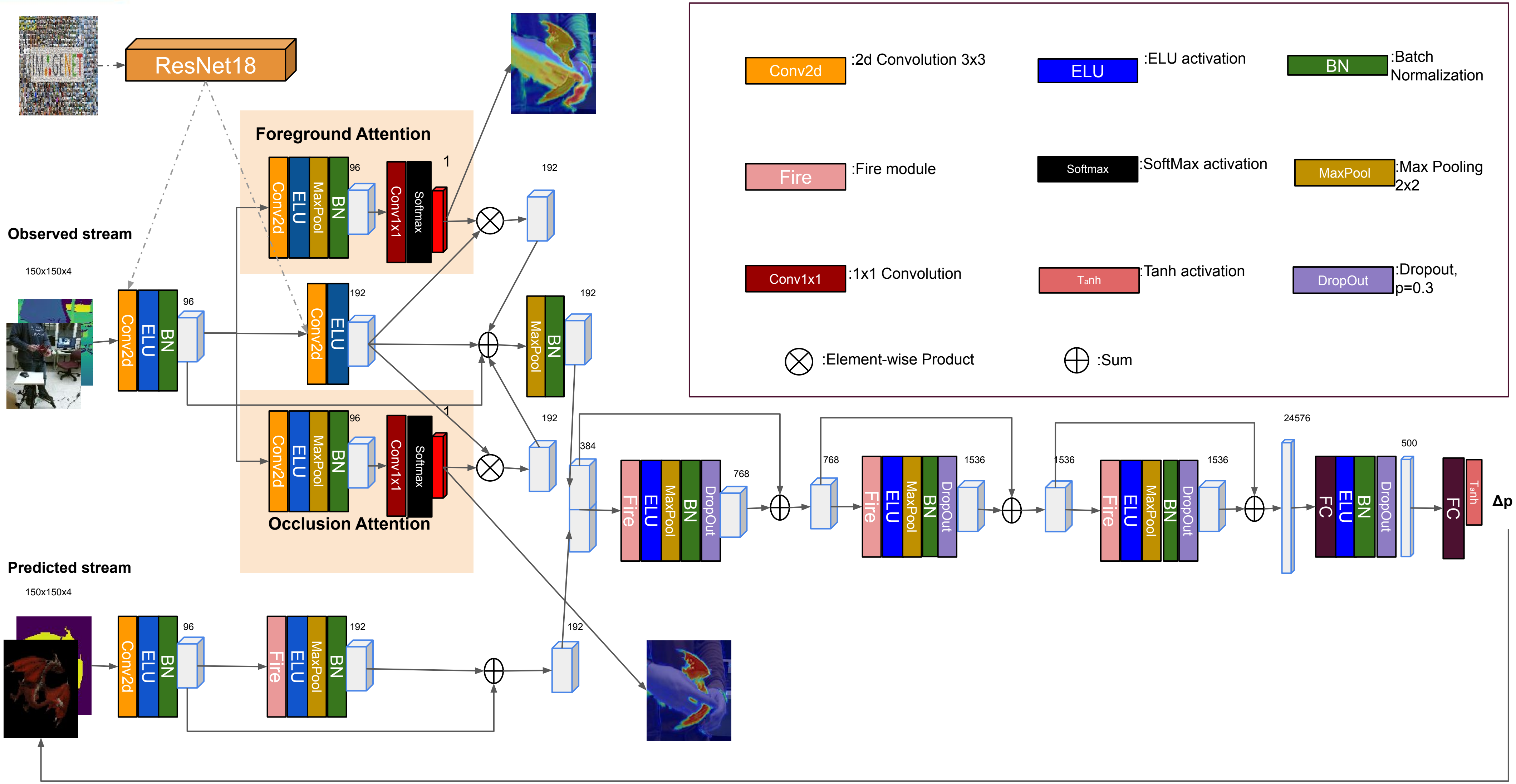


	Translational Error(mm)	Rotational Error(degrees)
Garon et al. [8]	34.38 ± 24.65	36.38 ± 36.31
Only occlusion	17.60 ± 10.74	37.10 ± 35.08
Hierarchical clutter & occlusion	**14.99 ± 9.89**	39.07 ± 33.22
Parallel clutter & occlusion	14.35 ± 10.21	**34.28 ± 29.81**

- Sharper peaks ⟶ Boosts accuracy

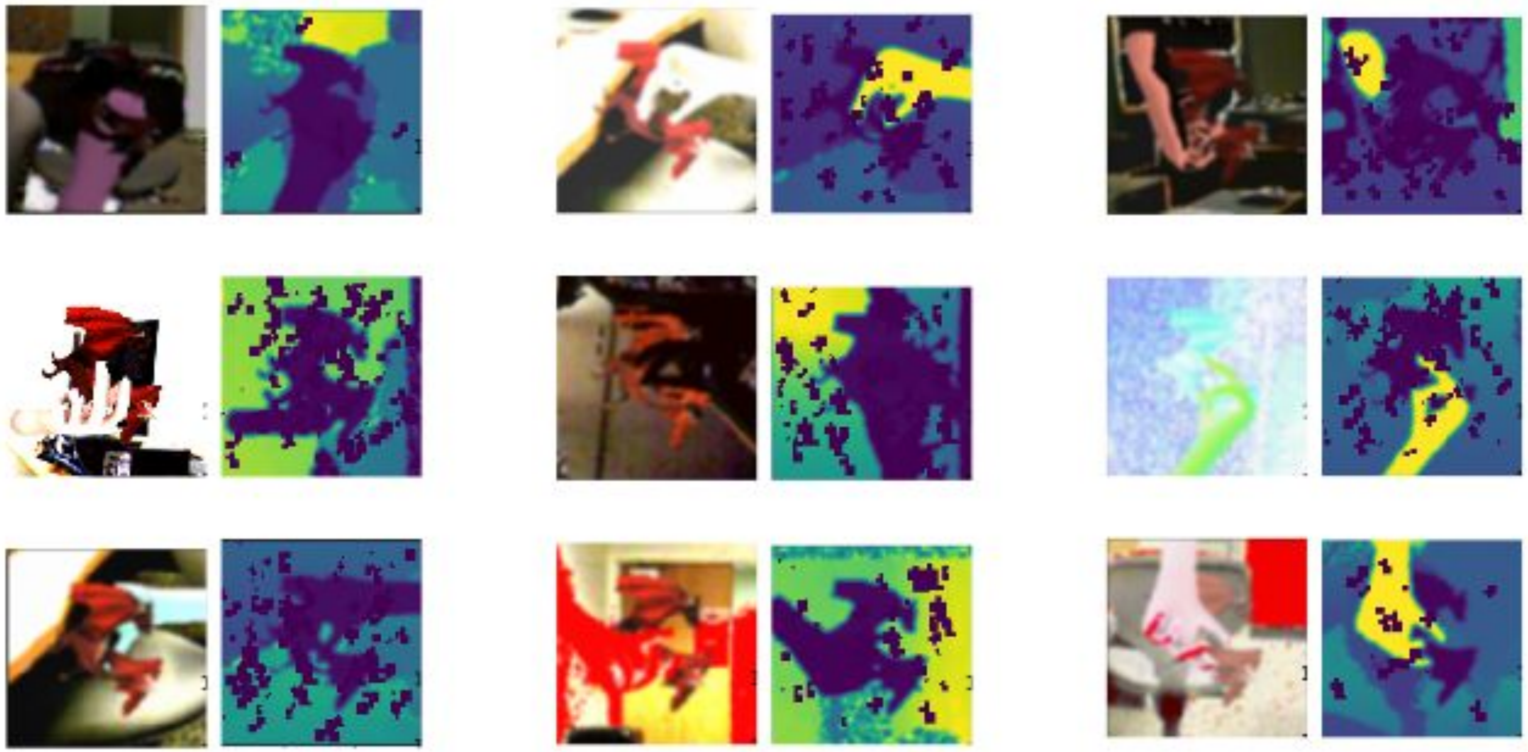
Our Overall Architecture:

- 20,000 training sample pairs
- Inference time: **40fps** (*real-time*)



Training data preprocessing:

- Pose sampling: **Golden spiral approach**[21]
- **Augmentation:** SUN3D[37] background, Occluder (*Partial & Total*), Gaussian Color & Depth Noise, Blur, Depth Holes, Color Jitter, Gamma Correction, Kinect Sensor noise modelling (Nguyen et al.[28])
- *Examples of Completely Augmented samples:*



- Recursive input standarization: **Welford's algorithm**[34]

Leopardi, P.C.: Distributing points on the sphere: partitions, separation, quadra-ture and energy. Ph.D. thesis, University of New South Wales, Sydney, Australia(2007)

Nguyen, C.V., Izadi, S., Lovell, D.: Modeling kinect sensor noise for improved3d reconstruction and tracking. In: 2012 Second International Conference on 3DImaging, Modeling, Processing, Visualization & Transmission. pp. 524–530. IEEE(2012)

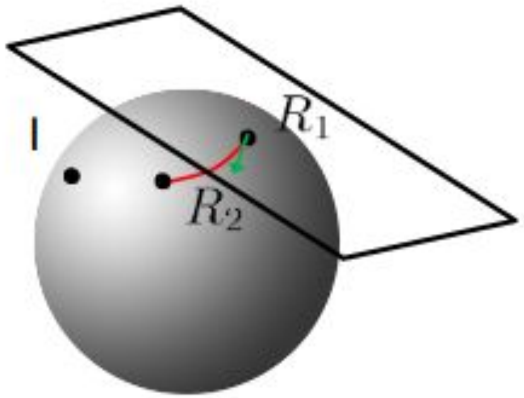
Xiao, J., Owens, A., Torralba, A.: Sun3d: A database of big spaces reconstructedusing sfm and object labels. In: Proc. IEEE Int. Conf. on Computer Vision (ICCV).pp. 1625–1632 (2013

Welford, B.: Note on a method for calculating corrected sums of squares and prod-ucts. Technometrics4(3), 419–420 (1962)

Garon, M., Lalonde, J.F.: Deep 6-dof tracking. IEEE transactions on visualizationand computer graphics23(11), 2410–2418 (2017)

Geodesic Rotational Loss

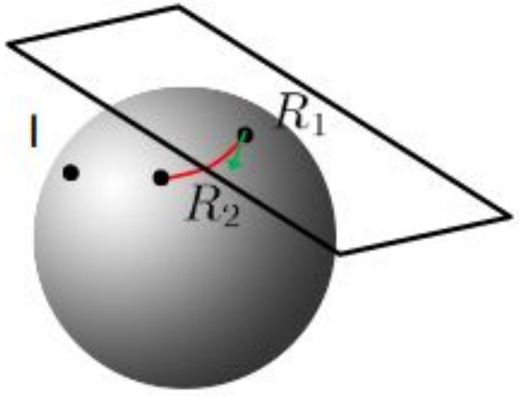
Garon et al.[8]: $L(\hat{\mathbf{p}}, \mathbf{p_{GT}}) = MSE(\hat{\mathbf{p}}, \mathbf{p}_{GT})$ with $\hat{\mathbf{p}}, \mathbf{p}_{GT} \in [-1, 1]^6$.



Riemannian metric

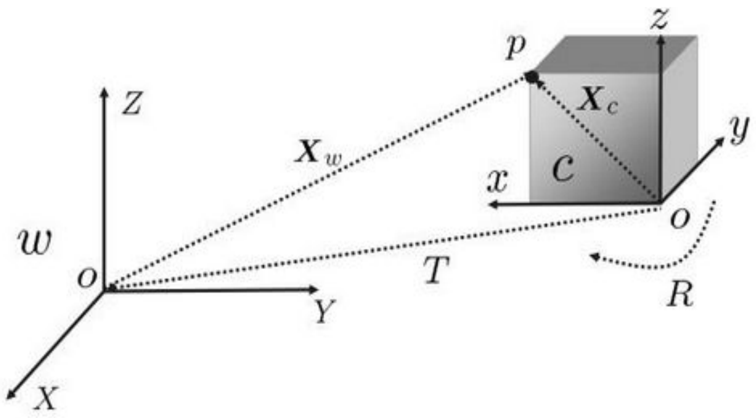
Geodesic Rotational Loss

Garon et al.[8]: $$L(\hat{\mathbf{p}}, \mathbf{p_{GT}}) = MSE(\hat{\mathbf{p}}, \mathbf{p}_{GT}) \quad \text{with } \hat{\mathbf{p}}, \mathbf{p}_{GT} \in [-1, 1]^6.$$

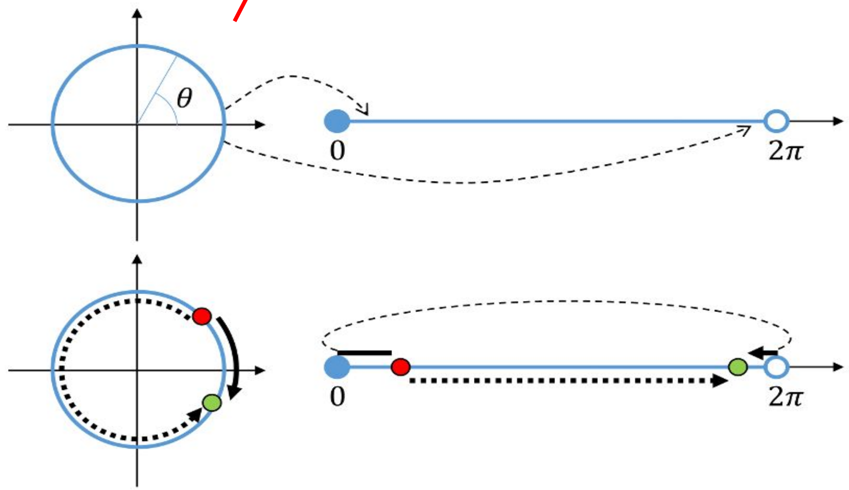


Riemannian metric

Translation $\longrightarrow$ Euclidean space



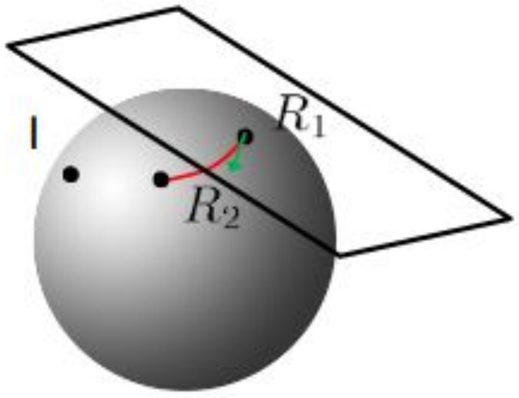
Rotation $\nrightarrow$ Euclidean space



Luca Ballan.Institute of Visual ComputingMetrics on SO(3) and Inverse Kinematics.http://lucaballan.altervista.org/pdfs/IK.pdf
Kris Hauser.Robotic Systems Book of Duke University.http://motion.pratt.duke.edu/RoboticSystems/3DRotations.html

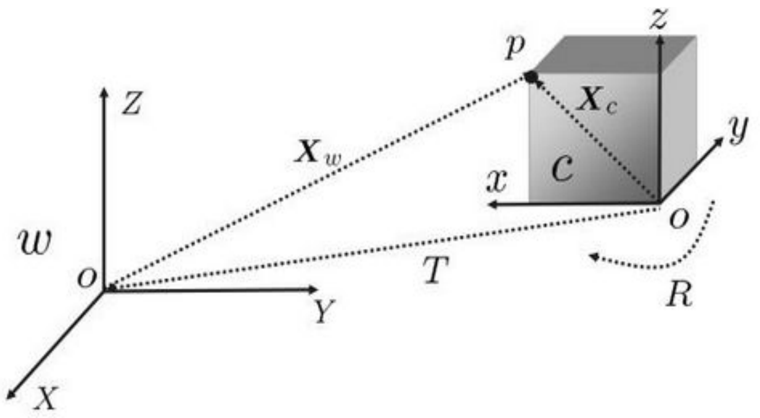
Geodesic Rotational Loss

Garon et al.[8]: $L(\hat{\mathbf{p}}, \mathbf{p}_{\mathbf{GT}}) = MSE(\hat{\mathbf{p}}, \mathbf{p}_{GT})$ with $\hat{\mathbf{p}}, \mathbf{p}_{GT} \in [-1, 1]^6$.

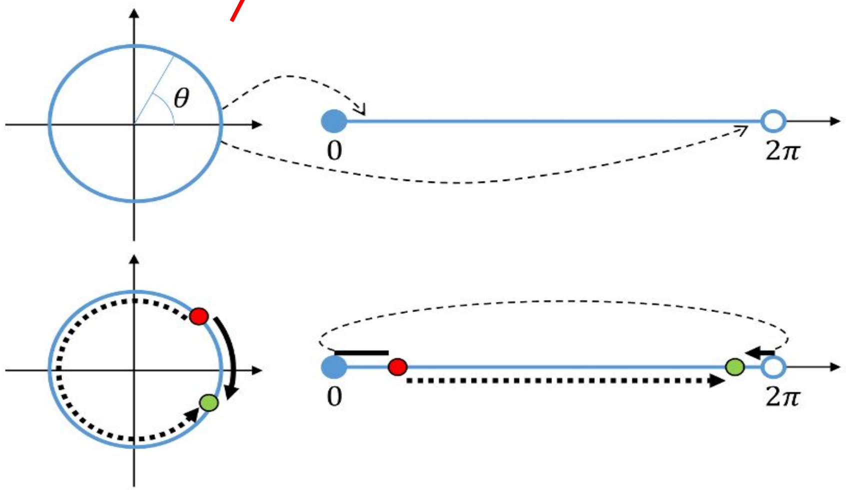


Riemannian metric

Translation $\longrightarrow$ Euclidean space



Rotation $\not\longrightarrow$ Euclidean space

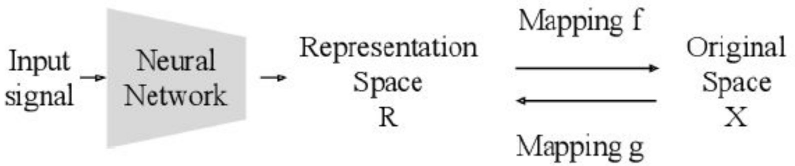


Geodesic (Riemannian) Rotational Loss function:

$$L_{Rot}(\Delta\hat{R}, \Delta R_{GT}) = d^{(Geod)}_{Rot}(\Delta\hat{R}, \Delta R_{GT}) = \arccos\left(\frac{\mathrm{Tr}\left(\Delta\hat{R}^T \Delta R_{GT}\right) - 1}{2}\right)$$

Luca Ballan.Institute of Visual ComputingMetrics on SO(3) and Inverse Kinematics.http://lucaballan.altervista.org/pdfs/IK.pdf
Garon, M., Laurendeau, D., Lalonde, J.F.: A framework for evaluating 6-dof objecttrackers. In: Proc. European Conf. on Computer Vision (ECCV). pp. 582–597(2018)
Kris Hauser Robotic Systems Book of Duke University http://motion.pratt.duke.edu/RoboticSystems/3DRotations.html

...+6D Continuous Rotation Representation (Zhou et al.[39])

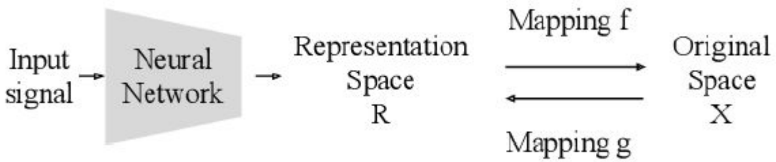


$$\Delta\mathbf{R_x} = N(\Delta\mathbf{r_x})$$
$$\Delta\mathbf{R_y} = N[\Delta\mathbf{r_y} - (\Delta\mathbf{R_x^T} \cdot \mathbf{r_y}) \cdot \boldsymbol{\Delta}\mathbf{R_x})]$$
$$\Delta\mathbf{R_z} = \Delta\mathbf{R_x} \times \Delta\mathbf{R_y}$$

where $\Delta\mathbf{R_{x/y/z}} \in \mathbb{R}^3$, $N(\cdot) = \frac{(\cdot)}{\|(\cdot)\|}$ is the normalization function.

Zhou, Y., Barnes, C., Lu, J., Yang, J., Li, H.: On the continuity of rotation rep-resentations in neural networks. In: Proc. IEEE Conf. on Computer Vision andPattern Recognition (CVPR). pp. 5745–5753 (2019)

...+6D Continuous Rotation Representation (Zhou et al.[39])

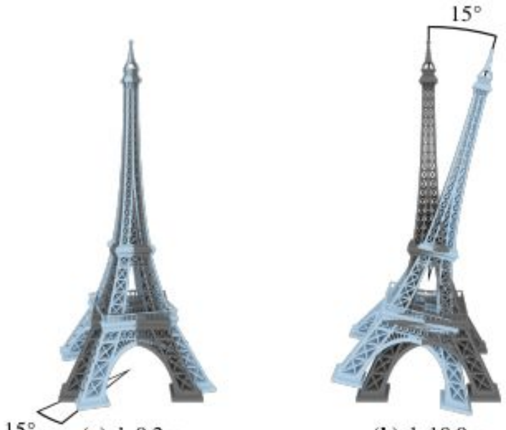


$$\Delta\mathbf{R_x} = N(\Delta\mathbf{r_x})$$
$$\Delta\mathbf{R_y} = N[\Delta\mathbf{r_y} - (\Delta\mathbf{R_x^T} \cdot \mathbf{r_y}) \cdot \mathbf{\Delta R_x})]$$
$$\Delta\mathbf{R_z} = \Delta\mathbf{R_x} \times \Delta\mathbf{R_y}$$

where $\Delta\mathbf{R_{x/y/z}} \in \mathbb{R}^3$, $N(\cdot) = \frac{(\cdot)}{\|(\cdot)\|}$ is the normalization function.

Zhou, Y., Barnes, C., Lu, J., Yang, J., Li, H.: On the continuity of rotation rep-resentations in neural networks. In: Proc. IEEE Conf. on Computer Vision andPattern Recognition (CVPR). pp. 5745–5753 (2019)

...+Rotational Anisotropy Weighting (Brégier et al.[2]):



(a) d=9.2m (b) d=18.9m

Inertia Tensor

$$\Lambda = \sqrt{\frac{1}{S}\sum_i \sigma_{\mathbf{a}_i,\mathbf{b}_i,\mathbf{c}_i}}$$

Brégier, R., Devernay, F., Leyrit, L., Crowley, J.L.: Defining the pose of any 3drigid object and an associated distance. Int. J. of Comp. Vision (IJCV)126(6),571–596 (2018)

Rotation Loss:

Garon et al.[8]:

$\mathbf{p} \in \mathbb{R}^6$

+

MSE

	Rotational Error(degrees)
Garon et al. [8]	36.38 ± 36.31
Rotational MSE	46.55 ± 40.88
Geod.	37.69 ± 35.39
Geod.+[39]	14.90 ± 21.76
Geod.+[39]+$\Lambda_{(G.S.)}$	$\mathbf{9.99 \pm 13.76}$

Rotation Loss:

Garon et al.[8]:

$\mathbf{p} \in \mathbb{R}^6$

+

MSE

$\longrightarrow R \in SO(3) \longrightarrow L_{Geod}$

	Rotational Error(degrees)
Garon et al. [8]	36.38 ± 36.31
Rotational MSE	46.55 ± 40.88
Geod.	37.69 ± 35.39
Geod.+[39]	14.90 ± 21.76
Geod.+[39]+$\Lambda_{(G.S.)}$	$\mathbf{9.99 \pm 13.76}$

Rotation Loss:

Garon et al.[8]:

$\mathbf{p} \in \mathbb{R}^6$

+

MSE

Ours:

$\mathbf{t} \in \mathbb{R}^3$

+

$\mathbf{r} \in \mathbb{R}^6 \longrightarrow R \in SO(3) \longrightarrow L_{Geod}$

	Rotational Error(degrees)
Garon et al. [8]	36.38 ± 36.31
Rotational MSE	46.55 ± 40.88
Geod.	37.69 ± 35.39
Geod.+[39]	14.90 ± 21.76
Geod.+[39]+$\Lambda_{(G.S.)}$	$\mathbf{9.99 \pm 13.76}$

Zhou, Y., Barnes, C., Lu, J., Yang, J., Li, H.: On the continuity of rotation rep-resentations in neural networks. In: Proc. IEEE Conf. on Computer Vision andPattern Recognition (CVPR). pp. 5745–5753 (2019)

Rotation Loss:

Λ

Garon et al.[8]:

$\mathbf{p} \in \mathbb{R}^6$

+

MSE

Ours:

$\mathbf{t} \in \mathbb{R}^3$

+

$\mathbf{r} \in \mathbb{R}^6 \longrightarrow R \in SO(3) \longrightarrow L_{Geod}$

	Rotational Error(degrees)
Garon et al. [8]	36.38 ± 36.31
Rotational MSE	46.55 ± 40.88
Geod.	37.69 ± 35.39
Geod.+[39]	14.90 ± 21.76
Geod.+[39]+$\Lambda_{(G.S.)}$	$\mathbf{9.99 \pm 13.76}$

Zhou, Y., Barnes, C., Lu, J., Yang, J., Li, H.: On the continuity of rotation rep-resentations in neural networks. In: Proc. IEEE Conf. on Computer Vision andPattern Recognition (CVPR). pp. 5745–5753 (2019)

Rotation Loss:

$$\Lambda \xrightarrow{\text{Gramm-Schmidt Orthonormalization}} \Lambda_{(G.S.)} \in SO(3)$$

Garon et al.[8]:

$\mathbf{p} \in \mathbb{R}^6$

+

MSE

Ours:

$\mathbf{t} \in \mathbb{R}^3$

+

$\mathbf{r} \in \mathbb{R}^6 \longrightarrow R \in SO(3) \longrightarrow \otimes \longrightarrow L_{Geod}$

	Rotational Error(degrees)
Garon et al. [8]	36.38 ± 36.31
Rotational MSE	46.55 ± 40.88
Geod.	37.69 ± 35.39
Geod.+[39]	14.90 ± 21.76
Geod.+[39]+$\Lambda_{(G.S.)}$	$\mathbf{9.99 \pm 13.76}$

Zhou, Y., Barnes, C., Lu, J., Yang, J., Li, H.: On the continuity of rotation rep-resentations in neural networks. In: Proc. IEEE Conf. on Computer Vision andPattern Recognition (CVPR). pp. 5745–5753 (2019)

…+ MultiTask weighting (Kendall et al.[20]) ⟶ Pose Tracking

$$\boxed{\mathbf{v} = [v_1, v_2]}$$

$$L_{Track}(\Delta\hat{\mathbb{P}}, \Delta\mathbb{P}) = e^{(-v_1)} \cdot MSE[(\Delta\hat{\mathbf{t}}, \Delta\mathbf{t})] + v_1 + v_2 + \\ + e^{(-v_2)} \cdot arcos\left(\frac{\mathrm{Tr}\left(\left(\Delta\hat{R} \cdot \hat{G}^* \cdot \Lambda_{(G.S.)}\right)^T \cdot \left(\Delta R \cdot \Lambda_{(G.S.)}\right)\right) - 1}{2}\right)$$

<u>Local Optima problem:</u> Weight warm-up by first minimizing a **LogCosh** loss

Overall Loss function:

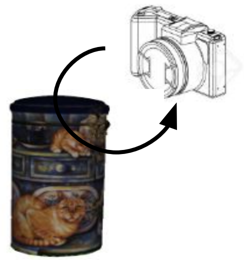
$$\boxed{\mathbf{s} = [s_1, s_2, s_3]}$$

$$Loss = e^{(-s_1)} \cdot L_{Track} + e^{(-s_2)} \cdot L_{Unoccl} + e^{(-s_3)} \cdot L_{Foregr} + s_1 + s_2 + s_3$$

Kendall, A., Gal, Y., Cipolla, R.: Multi-task learning using uncertainty to weighlosses for scene geometry and semantics. In: Proc. IEEE Conf. on Computer Visionand Pattern Recognition (CVPR). pp. 7482–7491 (2018)

Object Symmetries: Cases

Continuous Rotational Symmetries:



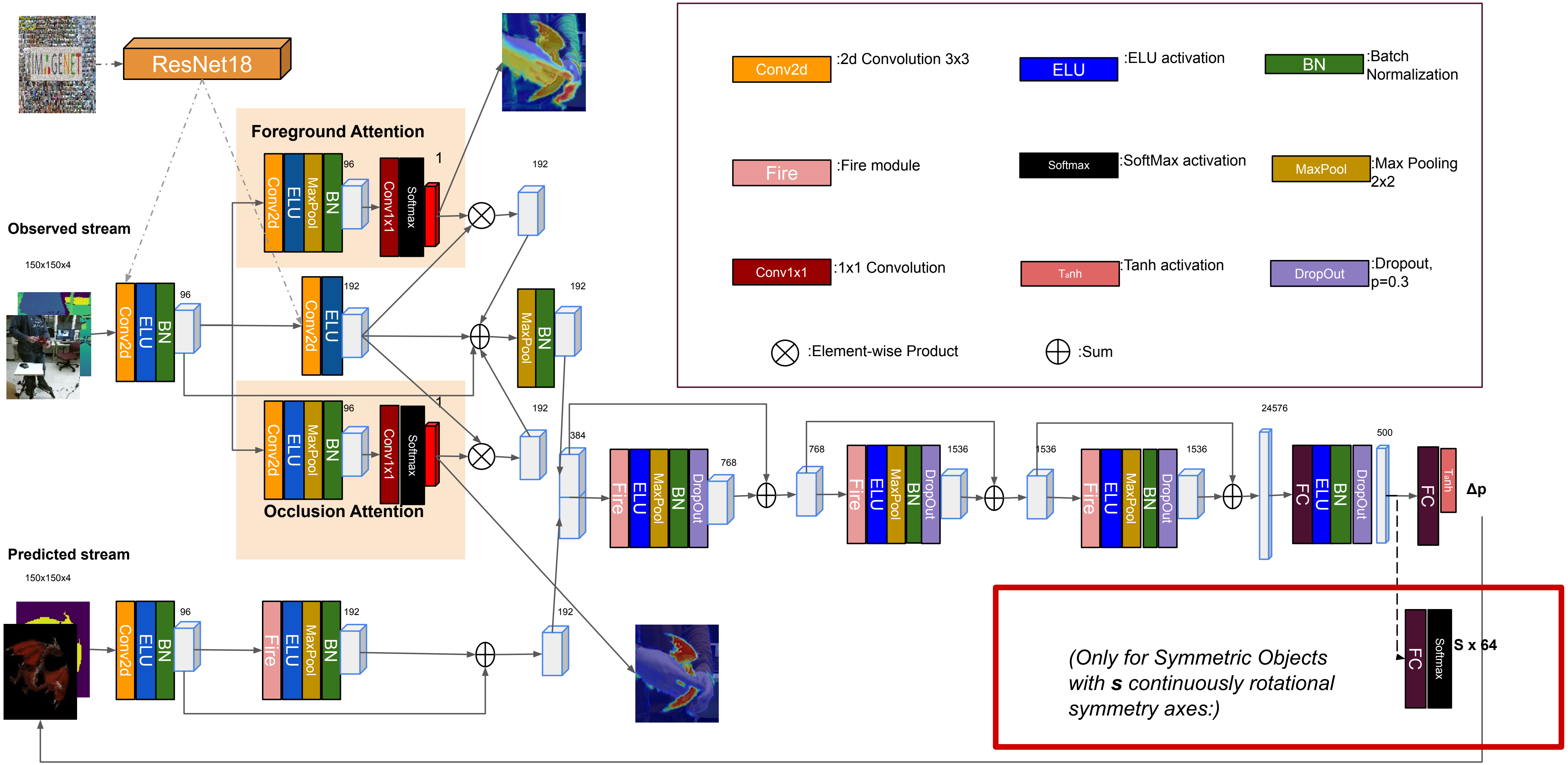
Object Symmetries: Cases

Continuous Rotational Symmetries:

Overall Loss function

(Symmetries' Handling incorporated):

Our Architecture *(Symmetries' Handling incorporated)*:



Object Symmetries: Cases

Continuous Rotational Symmetries:

Overall Loss function
(Symmetries' Handling incorporated):

Optimal symmetry parameter selection
out of a batch

$$Loss^{(Symm)} = Loss + e^{(-s_4)}\Big(\frac{1}{B}\sum_{b=1}^{B}\frac{1}{\xi_b}\Big) + s_4, \text{ with}$$

Object Symmetries: Cases

Continuous Rotational Symmetries:

Overall Loss function
(Symmetries' Handling incorporated):

Optimal symmetry parameter selection
out of a batch

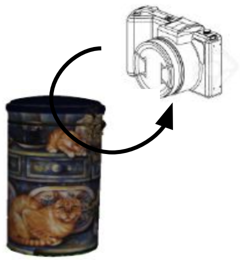
$$Loss^{(Symm)} = Loss + e^{(-s_4)}\left(\frac{1}{B}\sum_{b=1}^{B}\frac{1}{\xi_b}\right) + s_4, \text{ with}$$

$$\xi_b = \frac{1}{B_2(B_2-1)}\sum_{j=1}^{B_2}\sum_{k\neq j} d_{Rot}^{(Geod)}(\hat{G}_k, \hat{G}_j)$$

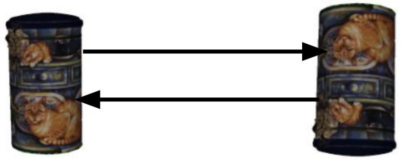
<u>Adversarial penalty</u> that encourages the **symmetry parameters** of the batch to be **as uniform as possible** by maximizng the rotational distances between them

Object Symmetries: Cases

Continuous Rotational Symmetries:



(Discrete) Reflective Symmetries:



Overall Loss function
(Symmetries' Handling incorporated):

Optimal symmetry parameter selection out of a batch

$$Loss^{(Symm)} = Loss + e^{(-s_4)}\left(\frac{1}{B}\sum_{b=1}^{B}\frac{1}{\xi_b}\right) + s_4, \text{ with}$$

$$\xi_b = \frac{1}{B_2(B_2-1)}\sum_{j=1}^{B_2}\sum_{k\neq j} d_{Rot}^{(Geod)}(\hat{G}_k, \hat{G}_j)$$

<u>Adversarial penalty</u> that encourages the **symmetry parameters** of the batch to be **as uniform as possible** by maximizng the rotational distances between them

Object Symmetries: Cases

Continuous Rotational Symmetries:

Overall Loss function
(Symmetries' Handling incorporated):

Optimal symmetry parameter selection
out of a batch

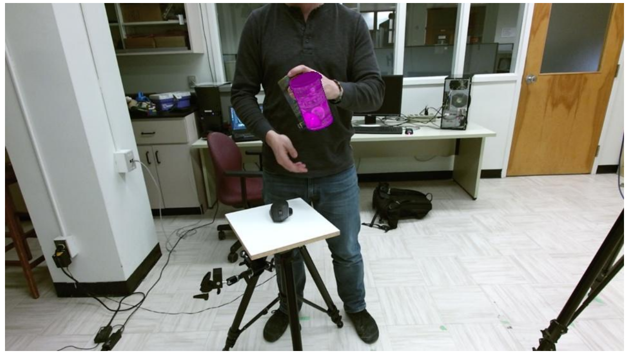
$$Loss^{(Symm)} = Loss + e^{(-s_4)}\left(\frac{1}{B}\sum_{b=1}^{B}\frac{1}{\xi_b}\right) + s_4, \text{ with}$$

$$\xi_b = \frac{1}{B_2(B_2-1)}\sum_{j=1}^{B_2}\sum_{k\neq j} d_{Rot}^{(Geod)}(\hat{G}_k, \hat{G}_j)$$

<u>Adversarial penalty</u> that encourages the **symmetry parameters** of the batch to be **as uniform as possible** by maximizng the distances between all of them

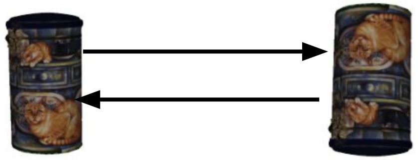
Frame no 127

Prediction no 127



Prediction no 126

(Discrete) Reflective Symmetries:



Object Symmetries: Cases

Continuous Rotational Symmetries:

Overall Loss function
(Symmetries' Handling incorporated):

Optimal symmetry parameter selection
out of a batch

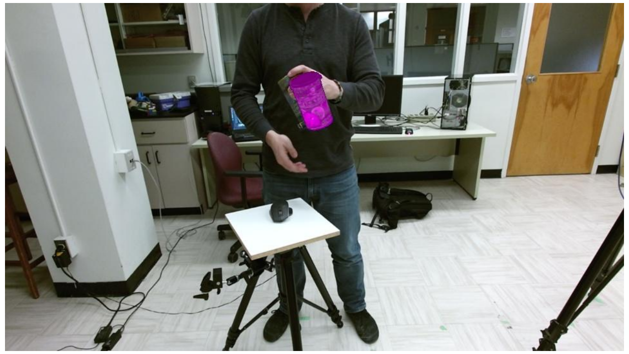
$$Loss^{(Symm)} = Loss + e^{(-s_4)}\Big(\frac{1}{B}\sum_{b=1}^{B}\frac{1}{\xi_b}\Big) + s_4, \text{ with}$$

$$\xi_b = \frac{1}{B_2(B_2-1)}\sum_{j=1}^{B_2}\sum_{k\neq j} d_{Rot}^{(Geod)}(\hat{G}_k, \hat{G}_j)$$

<u>Adversarial penalty</u> that encourages the **symmetry parameters** of the batch to be **as uniform as possible** by maximizng the distances between all of them

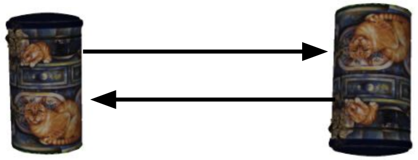
Frame no 127

Prediction no 127



Prediction no 126

(Discrete) Reflective Symmetries:



<u>Heuristic Algorithm:</u>

for *every Rotation estimation* $\hat{R}(t)$
do
  if $d_{Rot}(\hat{R}(t), \hat{R}(t-1)) \geq \left[\frac{360^o}{N_{DiscrSymm}} - th\right]$ **then**
    $\hat{R}(t-1) \rightarrow \hat{R}(t)$
  else
    $\hat{R}(t)$ is a valid estimation
  end
end

Dataset & Evaluation Metrics:

Object	Attributes				
	Size	Symmetry	Shape	Texture	Distinctive parts
Dragon	Medium	No	Complex	Rich	Yes
Cookie Jar	Medium	Rotoreflective	Simple	Poor and Repetitive	No
Dog	Medium	No	Complex	Almost None	Yes
Lego	Small	No	Complex	Rich and Repetitive	No
Watering Can	Big	No	Simple	Poor	Yes

Scenaria:

- 75% Horizontal Static Occlusion
- 75% Vertical Static Occlusion
- Translation Only
- Rotation Only
- Full Interaction
- Hard Interaction

3D Translational Error: (mm)

$$\delta_{\mathbf{t}}(\hat{\mathbf{t}}, \mathbf{t}_{GT}) = \left\|\hat{\mathbf{t}} - \mathbf{t}_{GT}\right\|_2$$

3D Rotational Error: (degrees)

$$\delta_R(\hat{R}, R_{GT}) = \arccos\left(\frac{\mathrm{Tr}\left(\hat{R}^T \cdot R_{GT}\right) - 1}{2}\right)$$

where $\mathrm{Tr}(\cdot)$ denotes the matrix trace.

Irrecoverable tracking fails:

When $\delta_{\mathbf{t}}(\hat{\mathbf{t}}, \mathbf{t}_{GT}) > 3cm$ or $\delta_R(\hat{R}, R_{GT}) > 20^o$

for more than7 consecutive frames.

Grey intervals: high occlusions

Green intervals: rapid motion

Garon, M., Laurendeau, D., Lalonde, J.F.: A framework for evaluating 6-dof objecttrackers. In: Proc. European Conf. on Computer Vision (ECCV). pp. 582–597(2018)

ECCV'20 ONLINE
23-28 AUGUST 2020

- Our tracker is generally on par or better with SoA across all objects and scenaria

Approach	75% Horizontal Occlusion Translational Error(mm)	75% Horizontal Occlusion Rotational Error(degrees)	75% Horizontal Occlusion Fails	75% Vertical Occlusion Translational Error(mm)	75% Vertical Occlusion Rotational(degrees)	75% Vertical Occlusion Fails
Garon et al.[8] ("Dragon")	16.02 ± **8.42**	18.35 ± 11.71	13	18.20 ± 11.81	14.66 ± 12.98	13
Ours("Dragon")	**12.68** ± 11.49	**13.00 ± 9.14**	**10**	**12.87 ± 10.49**	**13.14 ± 8.85**	**8**
Garon et al.[8]("Cookie Jar")	21.27 ± 9.74	21.90 ± 13.97	17	20.77 ± **6.88**	24.86 ± 13.64	20
Ours("Cookie Jar")	9.51 ± 4.17	15.48 ± 9.50	15	20.97 ± 7.32	16.14 ± 10.06	15
Ours+Symm.("Cookie Jar")	**6.37 ± 2.14**	**7.22 ± 3.97**	**11**	**19.01** ± 7.53	**13.00 ± 7.49**	**14**
Garon et al.[8]("Dog")	37.96 ± 23.39	47.94 ± 31.55	**21**	**32.84** ± 34.07	22.44 ± 13.60	21
Ours("Dog")	**24.43 ± 18.92**	**17.24 ± 12.41**	25	36.53 ± **22.39**	**12.67 ± 7.95**	**20**
Garon et al.[8]("Lego")	**68.25** ± 46.97	40.04 ± 47.37	28	40.04 ± 47.37	35.30 ± 31.32	20
Ours("Lego")	72.04 ± **34.10**	**18.41 ± 13.84**	**28**	**12.92 ± 5.73**	**12.92 ± 9.02**	**20**
Garon et al.[8]("Watering Can")	21.59 ± 11.32	23.99 ± **16.95**	14	32.76 ± 24.12	26.74 ± 19.05	18
Ours("Watering Can")	**20.71 ± 10.24**	**17.00** ± 18.99	**13**	**17.66 ± 17.95**	**13.46 ± 10.43**	**12**

Approach	Translation Interaction Translational Error(mm)	Translation Interaction Rotational Error(degrees)	Translation Interaction Fails	Rotation Interaction Translational Error(mm)	Rotation Interaction Rotational(degrees)	Rotation Interaction Fails
Garon et al.[8] ("Dragon")	41.60 ± 39.92	11.55 ± 15.58	15	23.86 ± 17.44	27.21 ± 22.40	15
Ours("Dragon")	**11.05 ± 8.20**	**3.55 ± 2.27**	**1**	**9.37 ± 6.07**	**7.86 ± 6.69**	**2**
Garon et al.[8]("Cookie Jar")	20.43 ± 25.44	17.19 ± 12.99	16	10.75 ± **5.89**	23.53 ± 18.85	19
Ours("Cookie Jar")	8.64 ± 8.23	8.31 ± 5.97	5	10.87 ± 8.14	20.55 ± 18.06	16
Ours+Symm.("Cookie Jar")	**8.09 ± 7.67**	**5.83 ± 5.50**	**3**	**9.98** ± 10.63	**13.84 ± 11.87**	**16**
Garon et al.[8]("Dog")	58.87 ± 71.86	16.42 ± 13.51	20	11.16 ± 10.28	**20.00** ± 21.31	17
Ours("Dog")	**21.64 ± 22.78**	**9.27 ± 8.03**	**14**	**10.68 ± 7.53**	20.07 ± **19.29**	**17**
Garon et al.[8]("Lego")	27.90 ± **23.53**	11.89 ± 18.50	29	16.42 ± 10.90	17.83 ± 15.90	32
Ours("Lego")	**22.66** ± 24.58	**9.08 ± 7.60**	**12**	**10.13 ± 6.79**	**7.22 ± 4.55**	**4**
Garon et al.[8]("Watering Can")	24.95 ± 42.91	13.26 ± 11.34	16	13.14 ± **8.99**	22.19 ± 25.93	15
Ours("Watering Can")	**24.30 ± 21.51**	**8.79 ± 6.35**	**16**	**12.22** ± 9.46	**18.66 ± 15.51**	**15**

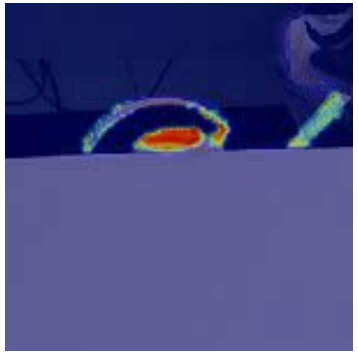
Apporach	Full Interaction Translational Error(mm)	Full Interaction Rotational Error(degrees)	Full Interaction Fails	Hard Interaction Translational Error(mm)	Hard Interaction Rotational(degrees)	Hard Interaction Fails
Garon et al.[8] ("Dragon")	35.23 ± 31.97	34.98 ± 29.46	18	34.38 ± 24.65	36.38 ± 36.31	17
Ours("Dragon")	**10.31 ± 8.66**	**6.40 ± 4.52**	**1**	**11.63 ± 8.79**	**8.31 ± 6.76**	**2**
Garon et al.[8]("Cookie Jar")	**13.06 ± 9.35**	31.78 ± 23.78	24	15.78 ± 10.43	24.29 ± 20.84	15
Ours("Cookie Jar")	17.03 ± 11.94	22.24 ± 20.86	21	15.29 ± 16.06	16.73 ± 14.79	11
Ours+Symm.("Cookie Jar")	14.63 ± 11.19	**15.71 ± 13.80**	**21**	**14.96 ± 9.06**	**15.00 ± 13.20**	**8**
Garon et al.[8]("Dog")	37.73 ± 42.32	**20.77 ± 19.66**	23	23.95 ± 38.86	24.38 ± 26.39	20
Ours("Dog")	**24.88 ± 35.85**	28.52 ± 25.38	**20**	**19.32 ± 15.97**	**19.72 ± 20.17**	**19**
Garon et al.[8]("Lego")	30.96 ± 31.44	22.10 ± 20.20	20	30.71 ± 42.62	36.38 ± 34.99	20
Ours("Lego")	**23.58 ± 27.73**	**11.80 ± 12.28**	**13**	**16.47 ± 12.95**	**14.29 ± 11.68**	**11**
Garon et al.[8]("Watering Can")	33.76 ± 37.62	40.16 ± 35.90	26	28.31 ± 19.49	23.04 ± 24.27	28
Ours("Watering Can")	**19.82 ± 19.98**	**28.76 ± 30.27**	**26**	**18.03 ± 14.99**	**19.57 ± 17.47**	**23**

‘75% Horizontal Occlusion’ scenario:

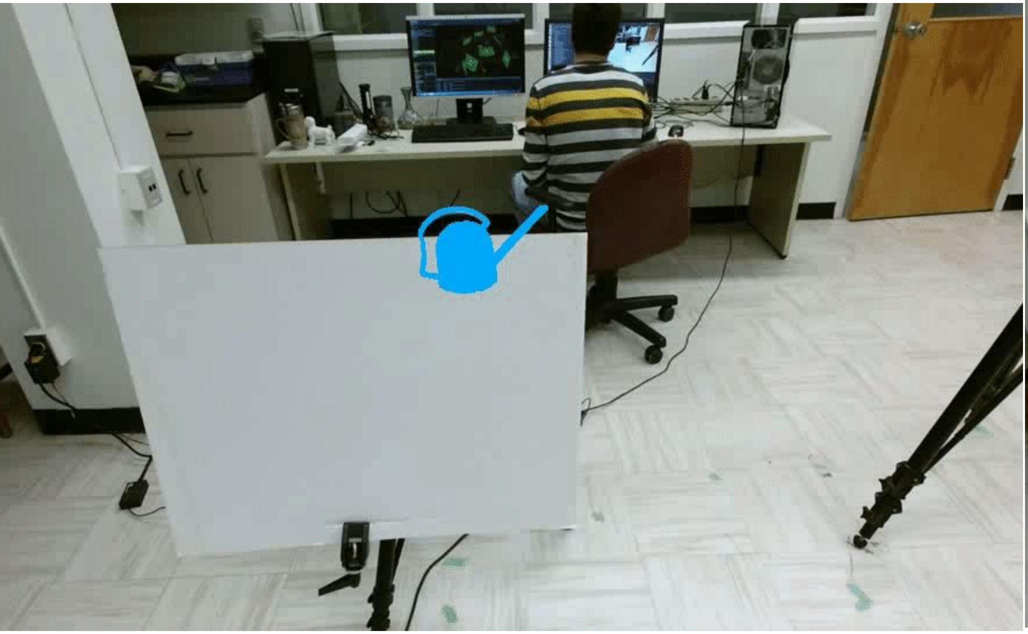
Qualitative Results: re-iterate every time the tracker fails irrecoverably

Foreground Attention:

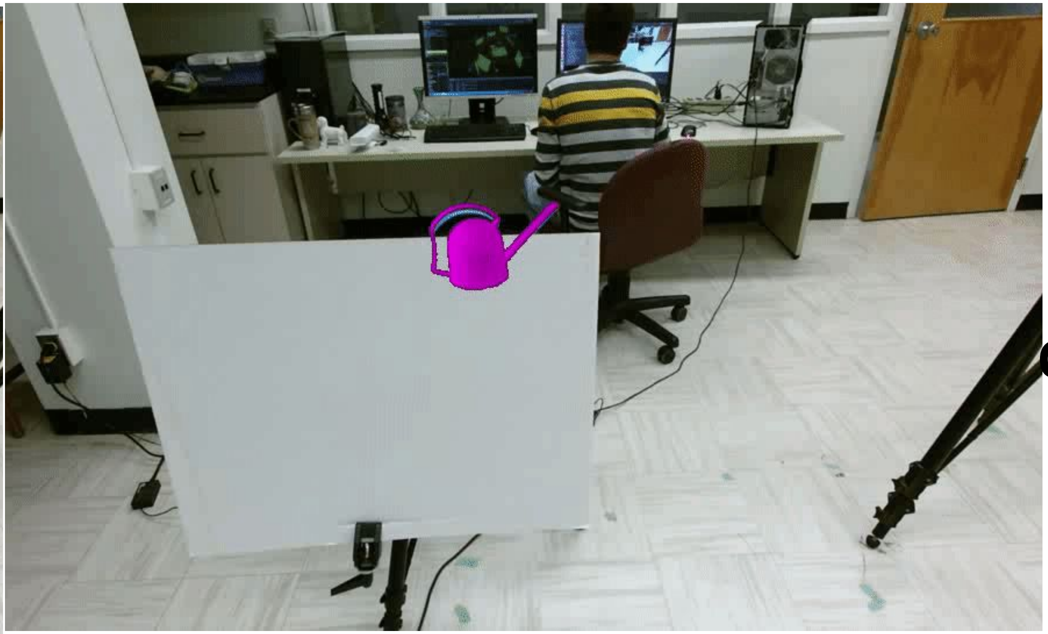
Occlusion Attention:



Garon et al.[8]

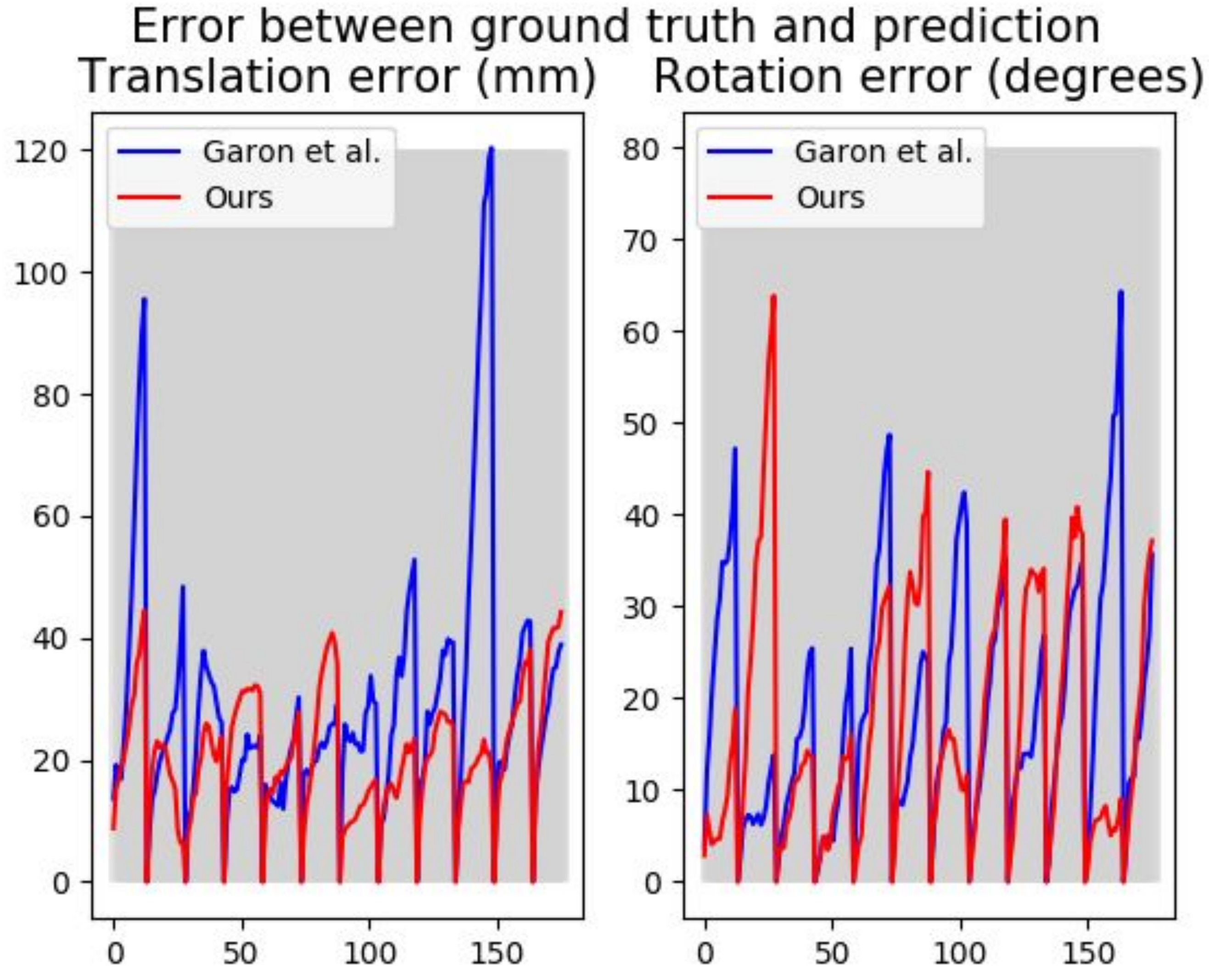


Ours



‘75% Horizontal Occlusion’ scenario:

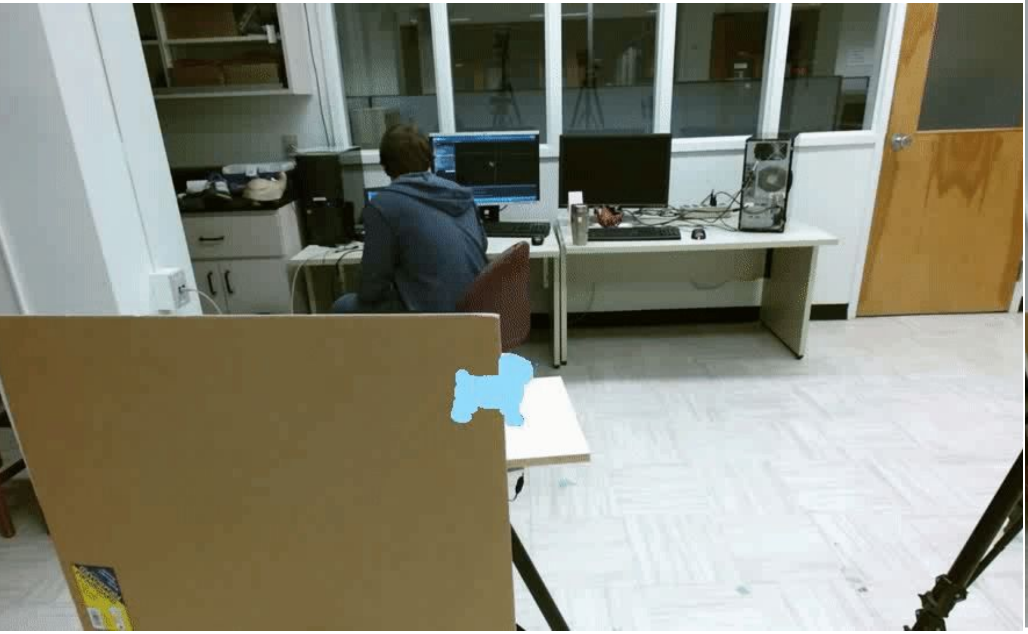
__Quantitative Results:__ re-iterate every 15 frames



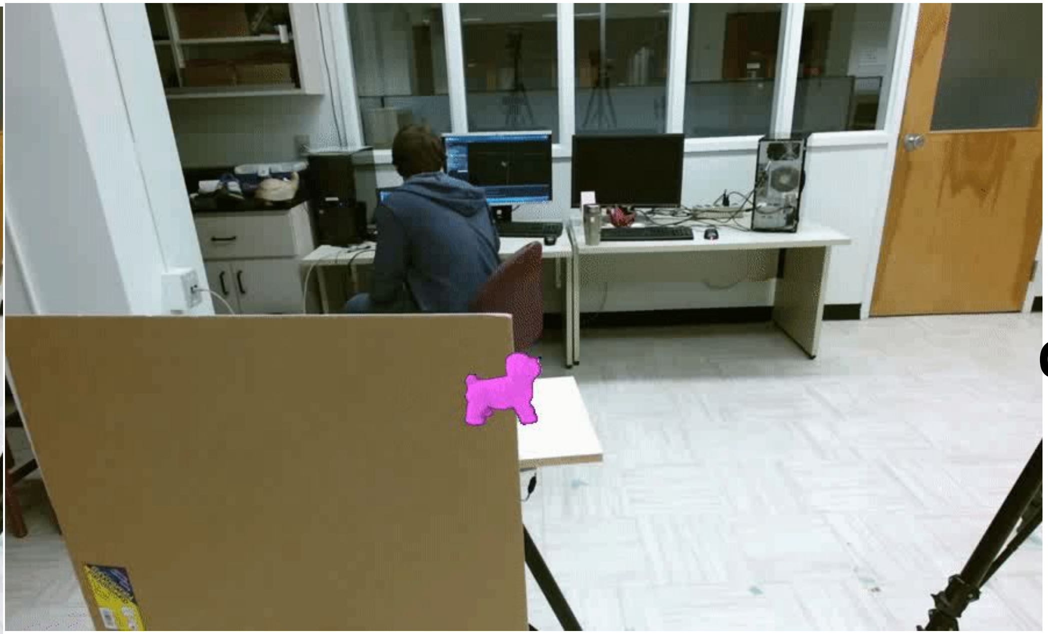
‘75% Vertical Occlusion’ scenario:

Qualitative Results: re-iterate every time the tracker fails irrecoverably

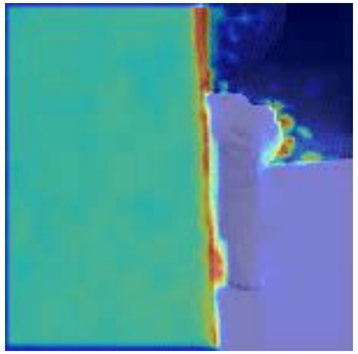
Garon et al.[8]



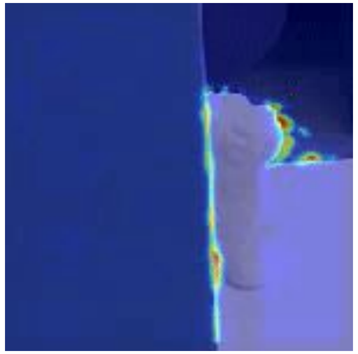
Ours



Foreground Attention:

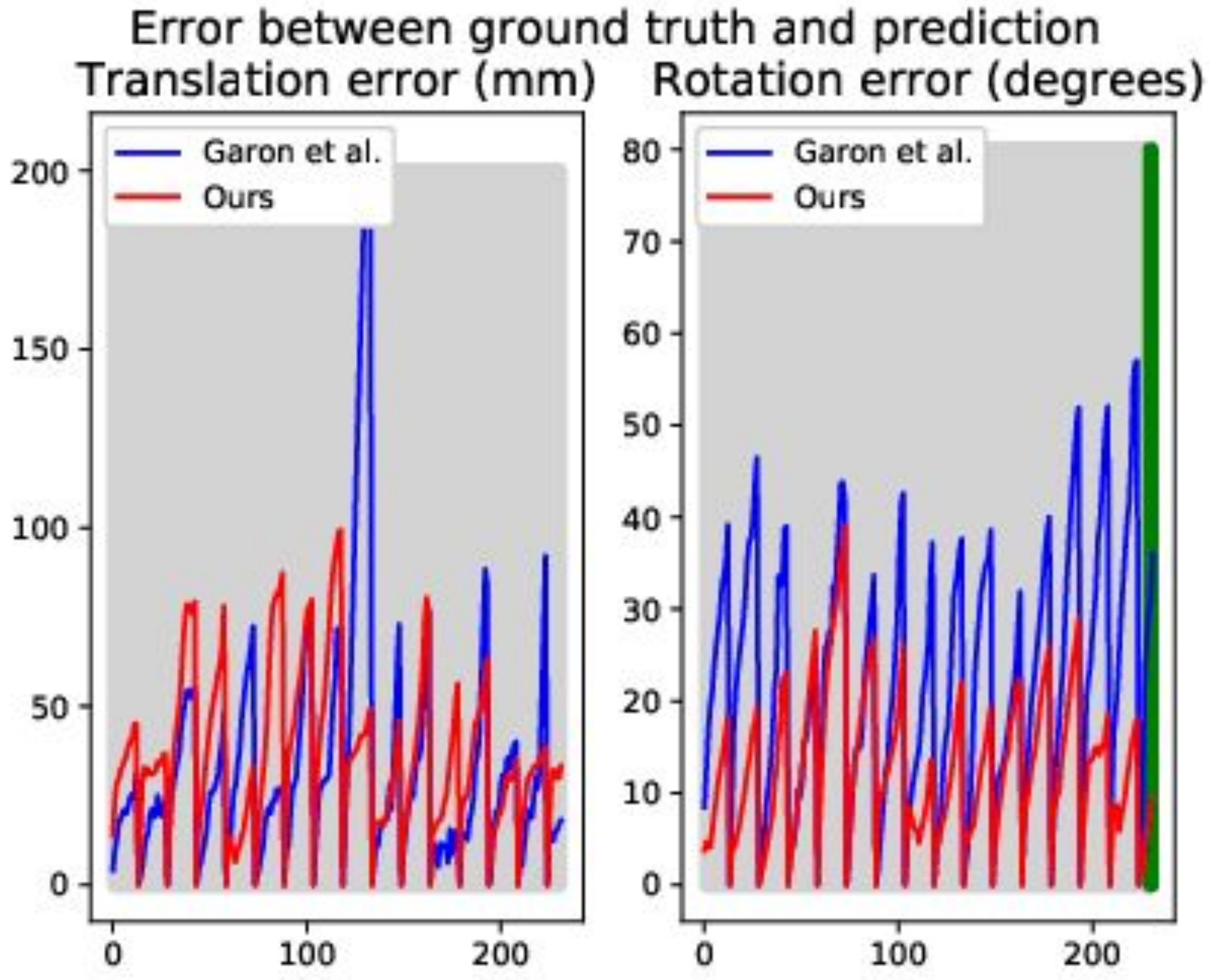


Occlusion Attention:



‘75% Vertical Occlusion’ scenario:

Quantitative Results: re-iterate every 15 frames



'Translation-Only Interaction' scenario:

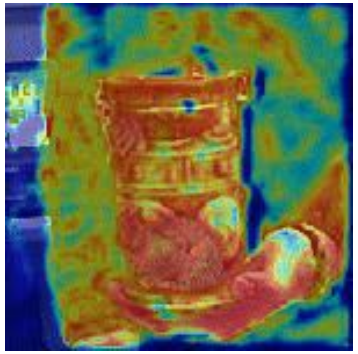
Qualitative Results: re-iterate every time the tracker fails irrecoverably

Garon et al.[8]

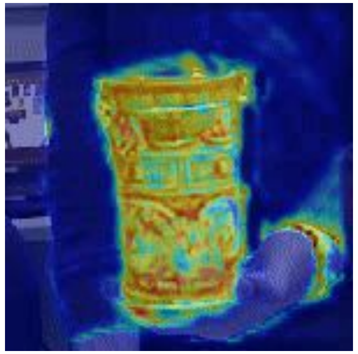
Ours

Ours+Symm

Foreground Attention:

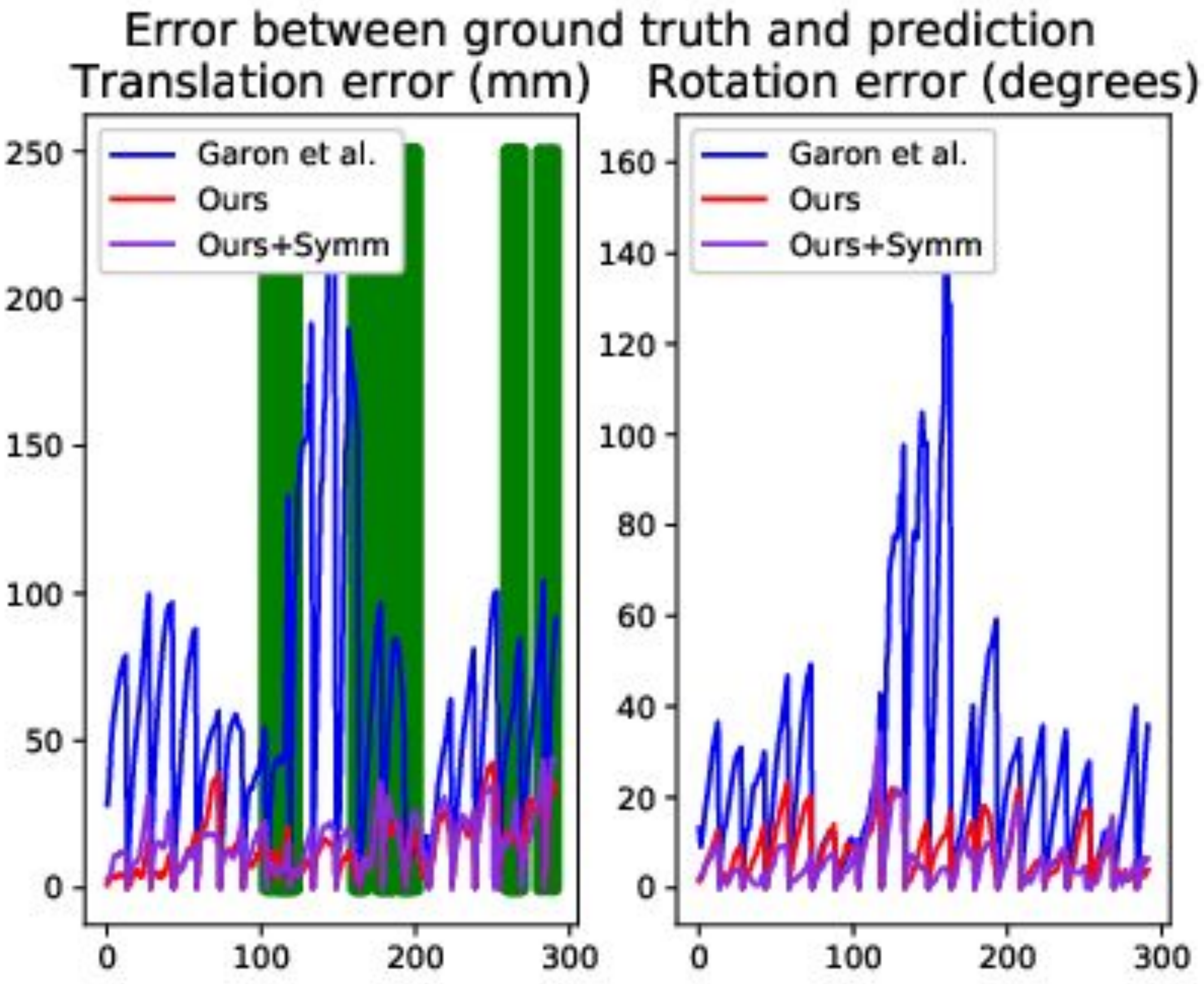


Occlusion Attention:



‘Translation-Only Interaction’ scenario:

Quantitative Results: re-iterate every 15 frames



‘Rotation-Only Interaction’ scenario:

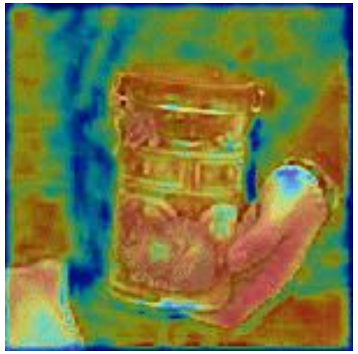
Qualitative Results: re-iterate every time the tracker fails irrecoverably

Garon et al.[8]

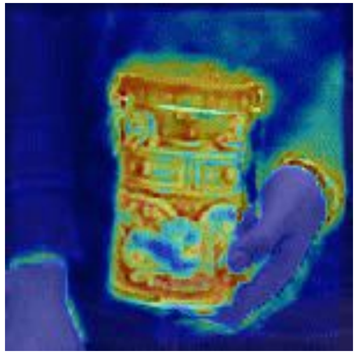
Ours

Ours+Symm

Foreground Attention:

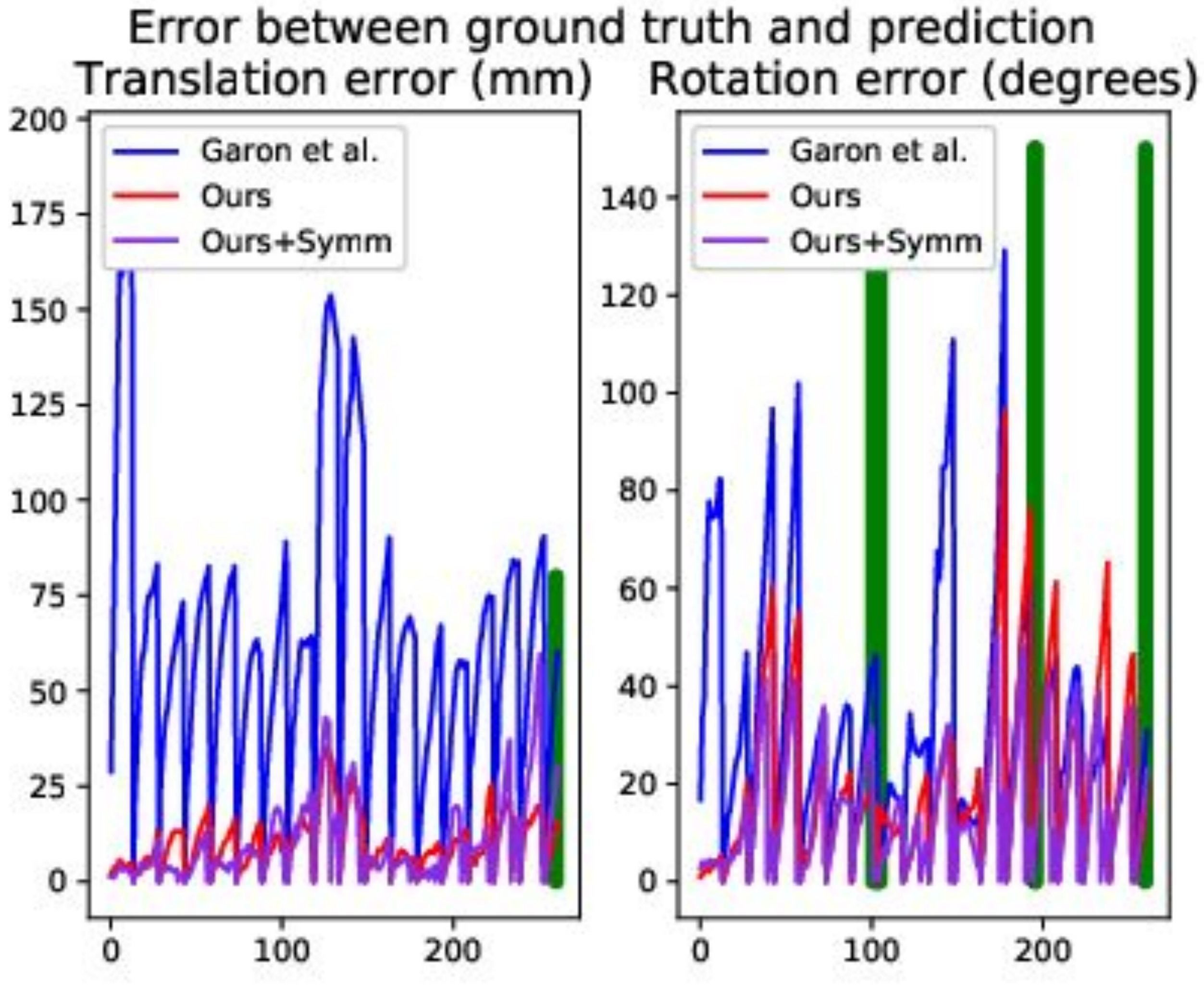


Occlusion Attention:



‘Rotation-Only Interaction’ scenario:

__Quantitative Results:__ re-iterate every 15 frames



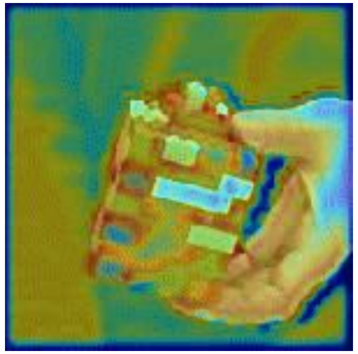
‘Full Interaction’ scenario:

<u>Qualitative Results:</u> re-iterate every time the tracker fails irrecoverably

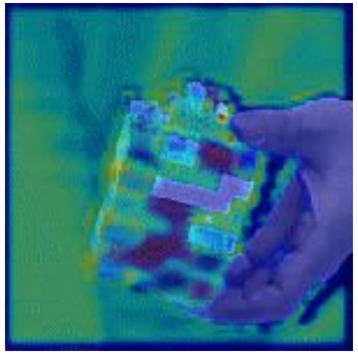
Garon et al.[8]

Ours

Foreground Attention:

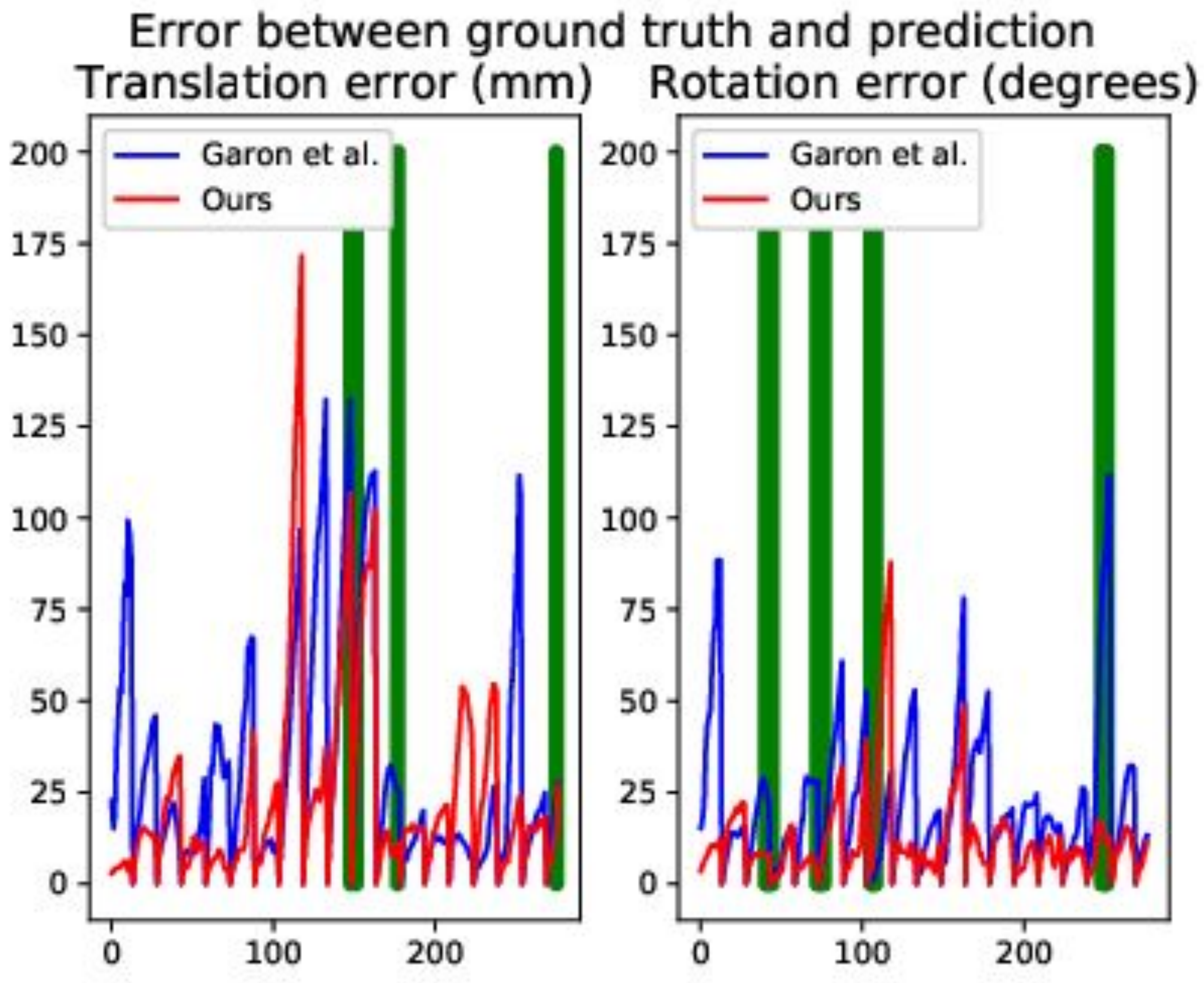


Occlusion Attention:



‘Full Interaction’ scenario:

<u>Quantitative Results:</u> re-iterate every 15 frames



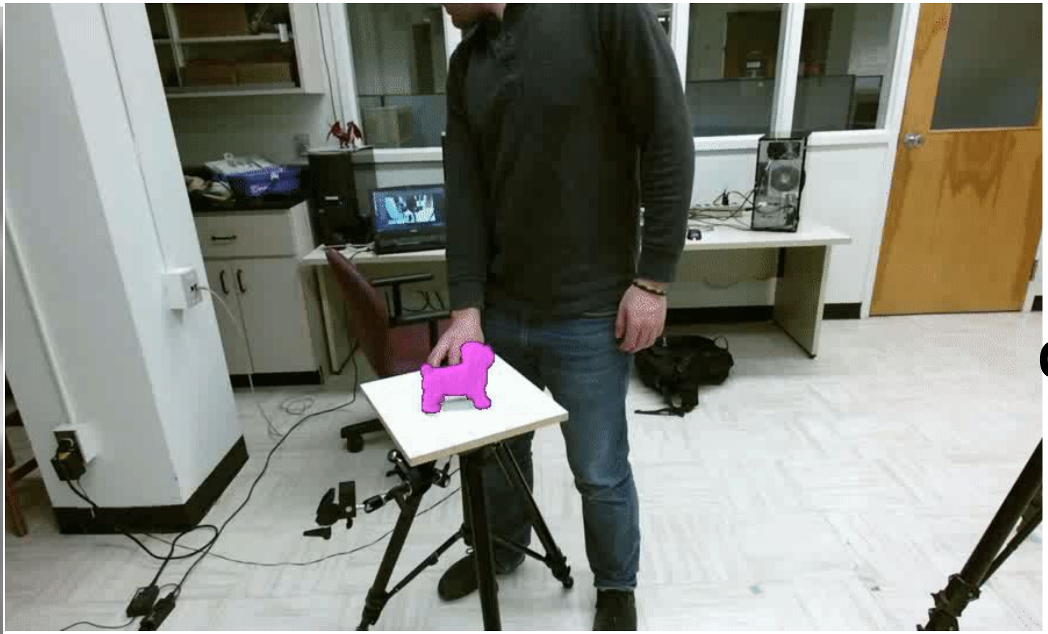
ECCV'20 ONLINE 23-28 AUGUST 2020

'Hard Interaction' scenario:

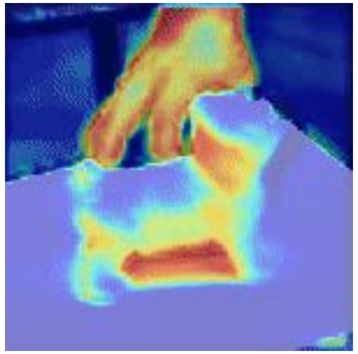
Qualitative Results: re-iterate every time the tracker fails irrecoverably

Garon et al.[8]

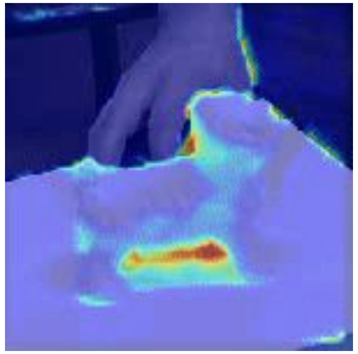
Ours



Foreground Attention:



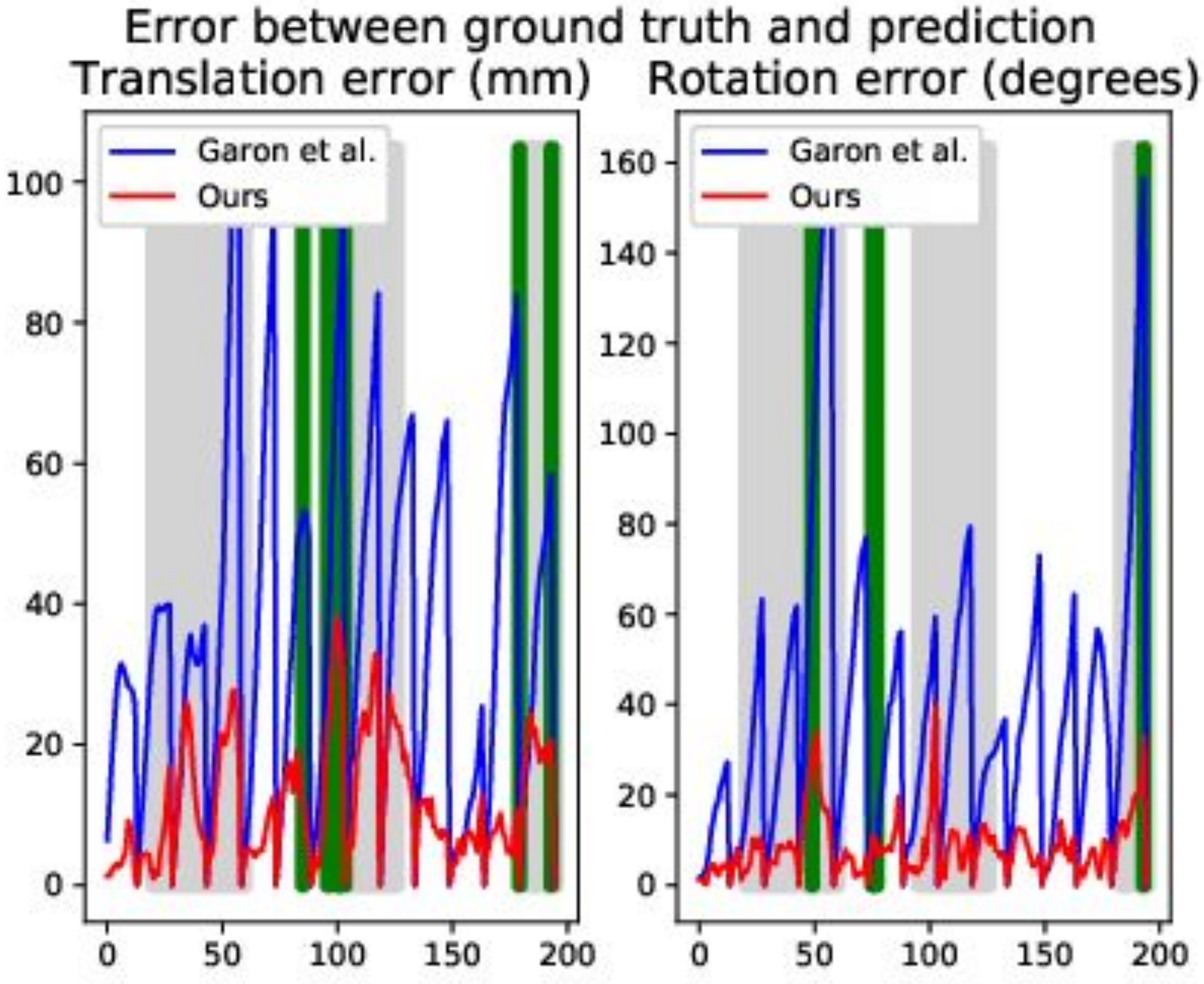
Occlusion Attention:



- Unsupervised learning of self-occlusion attention patterns
- Implicit learning of intuitive attentional regions of interest
- Visual tradeoff between the two parallel attention modules

‘Hard Interaction’ scenario:

Qualitative Results: re-iterate every time the tracker fails irrecoverably



Thank you!