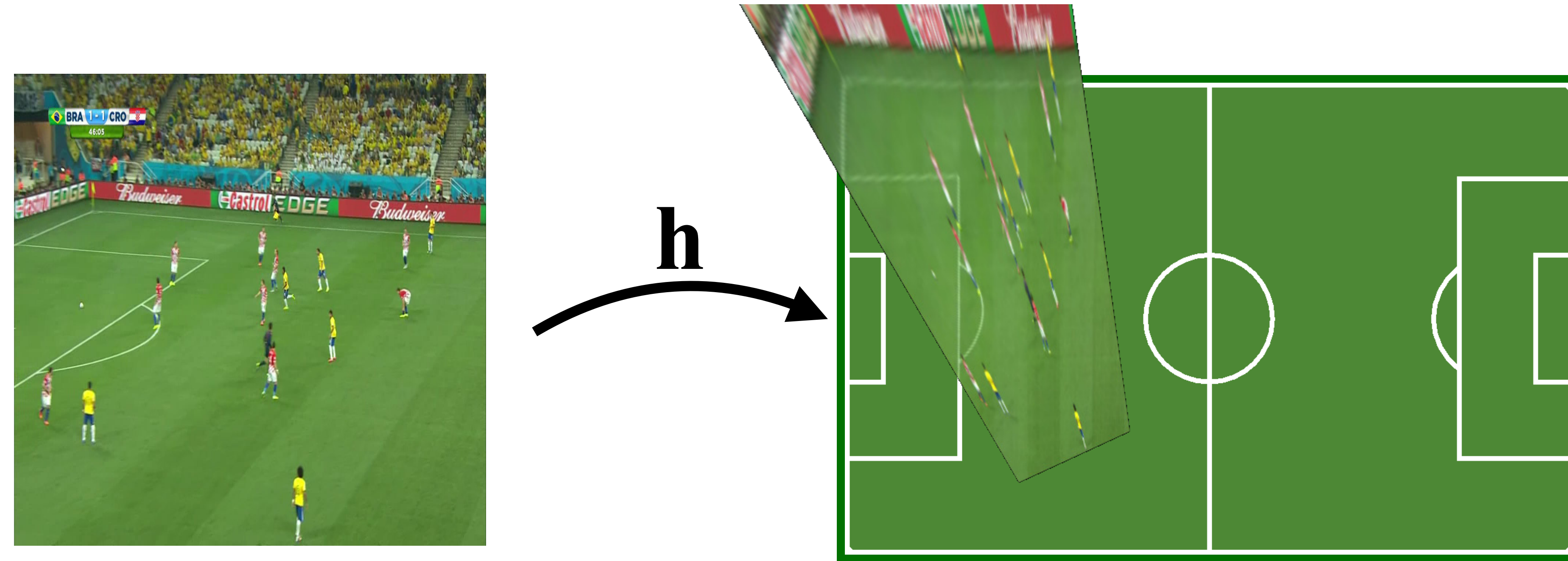
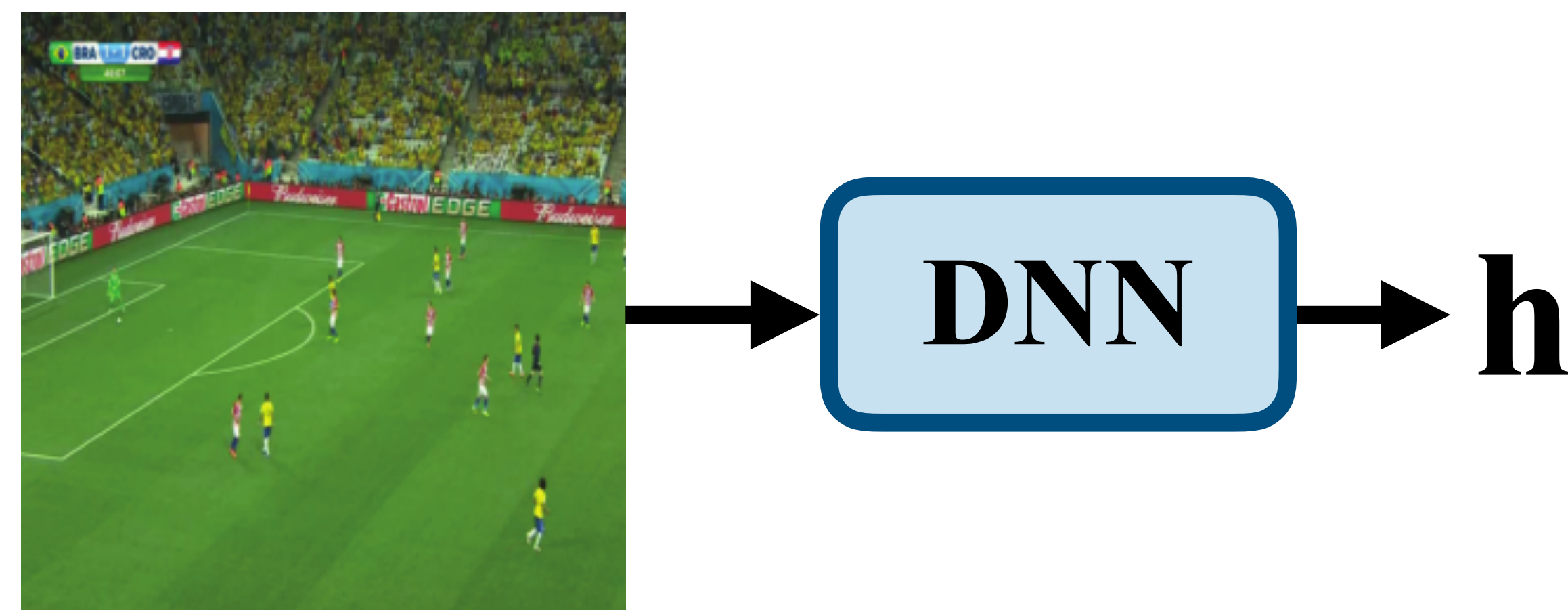


Problem

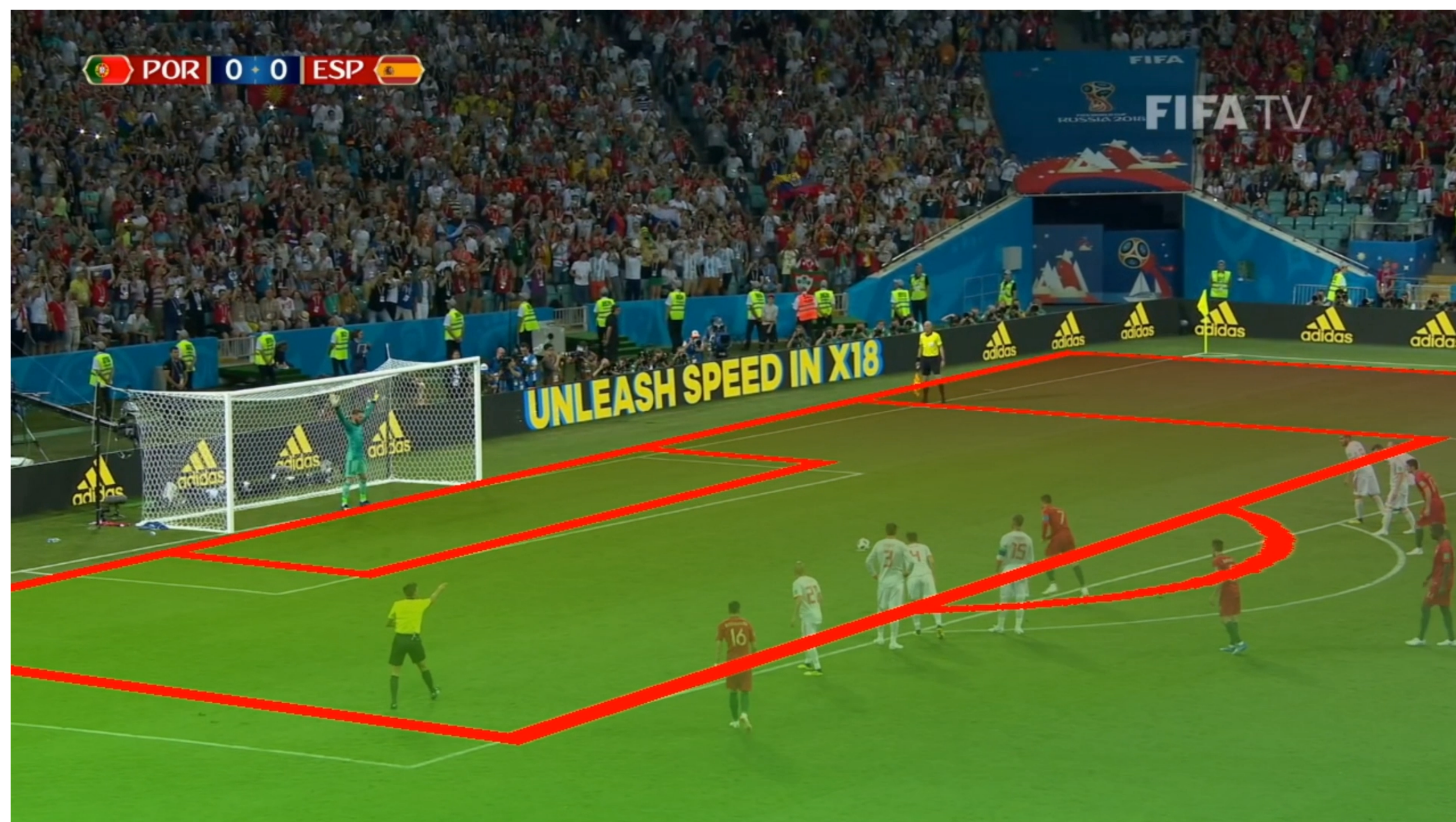
- Estimate the homography which can accurately maps a video frame to a sports field template.



How's feedforward network doing?

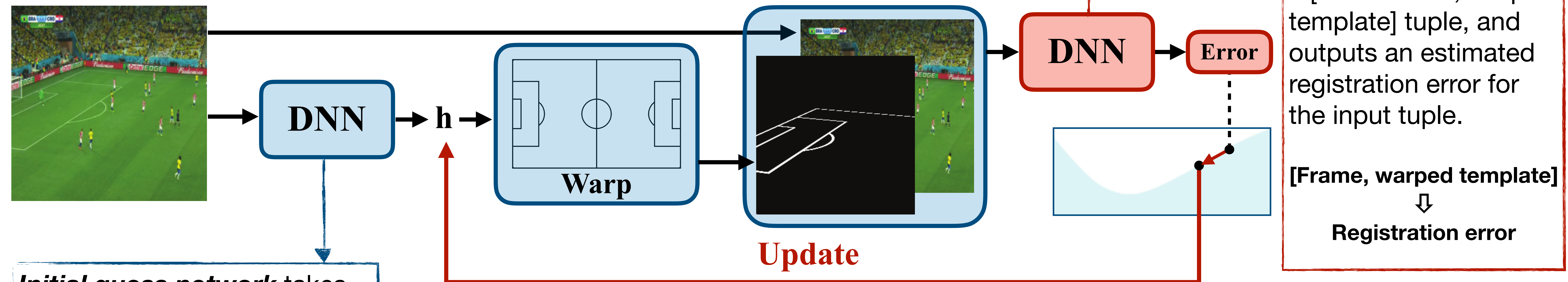


- We train a ResNet18 to regress to the homography, but it cannot estimate accurate homography.



- An example output of a trained ResNet, the red outline is the projected template.

Our method

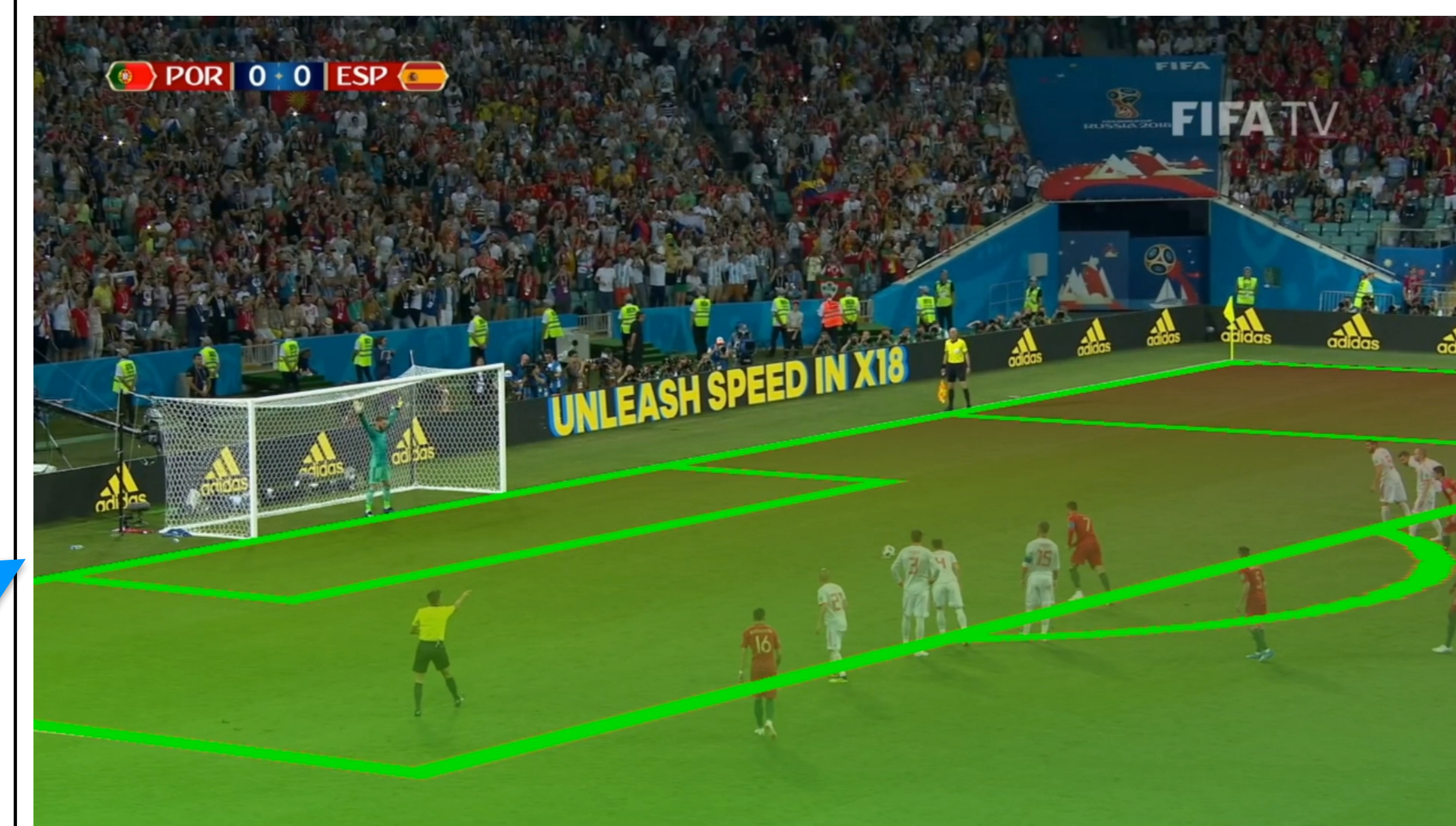


Initial guess network takes in a video frame, and outputs a homography which maps the frame to the template.

Frame \Rightarrow Homography

- Feed the input frame to the **initial guess network** to obtain a rough homography h .
- Warp the template using the estimated homography h .
- Concatenate the input frame with the warped template, and feed it to the **error network** to obtain the estimated registration error E .
- Back-propagate the error E to the estimated homography h , and obtain the gradients.
- Update homography h according to the optimizer and learning rate.
- Terminate if run out of budget, otherwise go to **step 2**.

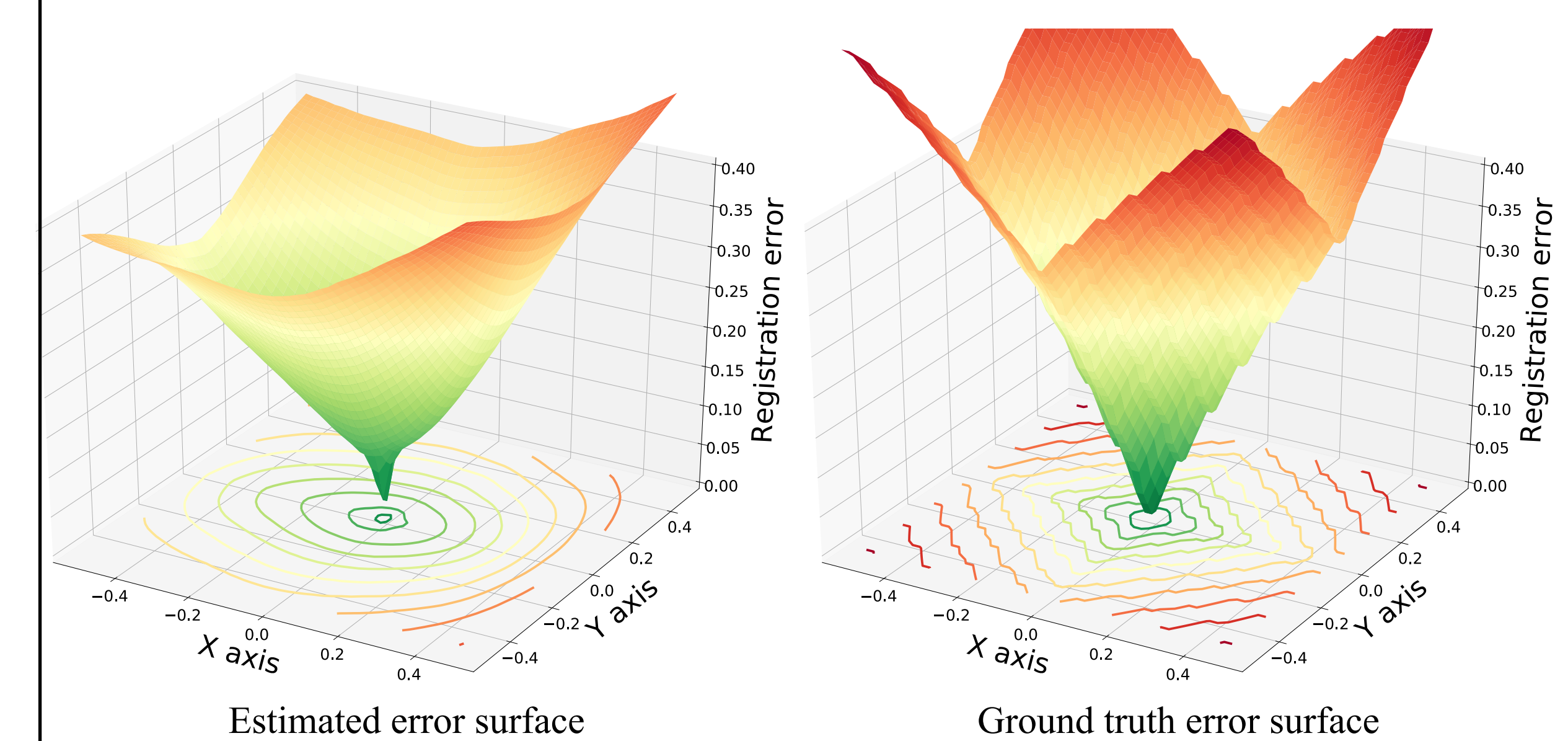
Results



			[19]	[42]	[6]	SFF	FFR	Ours
World Cup	IoU _{whole}	mean	83	—	89.2	83.9	84.0	89.8
		median	—	—	91.0	85.7	86.2	92.9
	IoU _{part}	mean	—	91.4	94.7	90.2	90.3	95.1
		median	—	92.7	96.2	91.9	92.1	96.7
Hockey	IoU _{whole}	mean	82 ¹	—	—	86.5	93.0	96.2
		median	—	—	—	87.3	94.0	97.0
	IoU _{part}	mean	—	—	—	90.4	96.0	97.6
		median	—	—	—	91.0	96.8	98.4

- The image on the left is the optimized result: video frame with warped template overlaid. The accuracy is much better than the feedforward method. Note that our model is trained on only 170 images, but can generalize to games from different time, stadiums, and camera poses and is able to provide accurate registrations.
- Quantitative results on the right. We achieved the state of the art performance on World Cup dataset and hockey dataset. More details please refer to our paper.
- Try our method at: <https://tinyurl.com/sportsfield>

Error surface



- Average estimated and ground truth error surface visualization for translation
- The estimated error surface resembles the ground truth one, and they share a common global minimum. This allows our method to work properly.