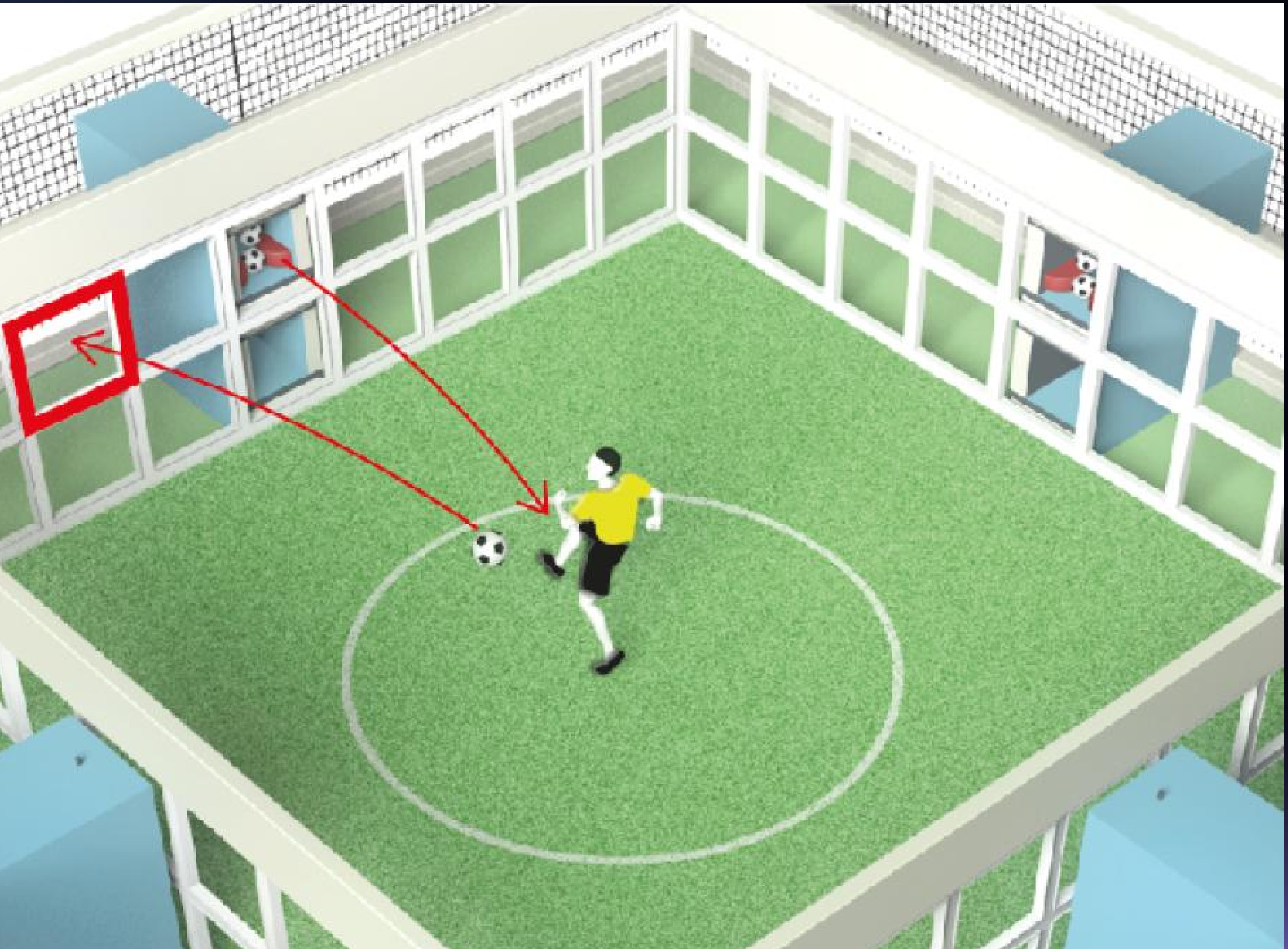
Anna Braun, Chihiro Tone, Raffael Rizzo

Democratising high-tech football training with Mixed Reality

Theoretical foundations

The Footbonaut



- from \$2.4 million to \$3.5 million (Source: statathlon.com, 2017)
- Borussia Dortmund, TSG 1899 Hoffenheim, Benfica Lissabon, RB Leipzig, Paris St. Germain … have the same or similar solutions
- Professional players and youth academy players train weekly with the Footbonaut

Agenda

01

Professional vs. Amateur Soccer Players

Physical, tactical and technical differences

02

Computer Vision in Sports

Extending human vision

03

XR in Sports

Merging virtual worlds with reality

04

Creating a MR Footbonaut

Democratizing the Footbonaut

Agenda

01

Professional vs. Amateur Soccer Players

Physical, tactical and technical differences

02

Computer Vision in Sports

Extending human vision

03

XR in Sports

Merging virtual worlds with reality

04

Creating a MR Footbonaut

Democratizing the Footbonaut

Research objectives

Der Footbonaut als Mess- und Informationssystem im Fußball

Eine explorative Untersuchung

Christian Saal[1] & Harald Fiedler[2]
Projekt CLIP
Projektleitung: Prof. Jochen Zinner[1] & Prof. Ralf Lanwehr[2]
Hochschule für Gesundheit & Sport, Technik & Kunst[1]
Business and Information Technology School[2]

- Find out whether higher-performing players achieve lower action times and higher hit rates than lower-performing players.

- As the Footbonaut was developed as a Measurement Information System for professional soccer in order to quantify speed of action and passing precision, the researchers used the Footbonaut to conduct this research.

Stein, Manuel, Janetzko, Halldor, Lamprecht, Andreas, Breitkreutz, Thorsten, Zimmermann, Philipp, Goldlucke, Bastian, Schreck, Tobias (Sport Analysis, 2018): Bring it to the Pitch: Combining Video and Movement Data to Enhance Team Sport Analysis, in: IEEE TVCG, 24 (2018), no. 1, p. 13-22.

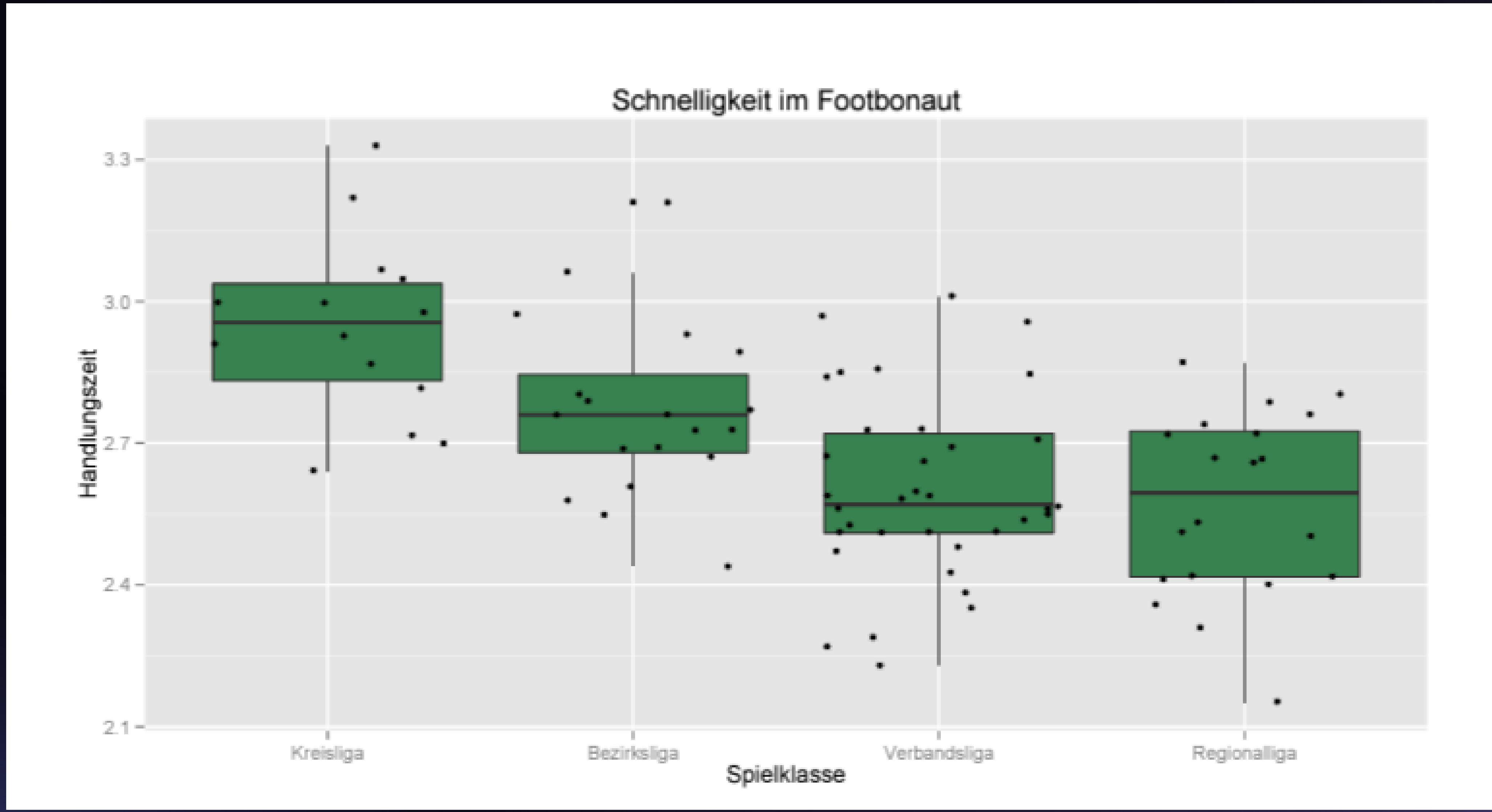
Experimental design

- The Footbonaut, a 14x14m artificial turf field enclosed by four walls containing 64 target squares, uses light barriers for action time and hit recordings.
- 88 junior soccer players from U16 teams participated in the study, sorted by league affiliation.
- Participants were required to kick balls launched from cannons into a target area indicated by visual and acoustic signals.
- After a warm-up, each player completed a trial with a randomly determined sequence of 32 balls.

Stein, Manuel, Janetzko, Halldor, Lamprecht, Andreas, Breitkreutz, Thorsten, Zimmermann, Philipp, Goldlucke, Bastian, Schreck, Tobias (Sport Analysis, 2018): Bring it to the Pitch: Combining Video and Movement Data to Enhance Team Sport Analysis, in: IEEE TVCG, 24 (2018), no. 1, p. 13-22.

Results

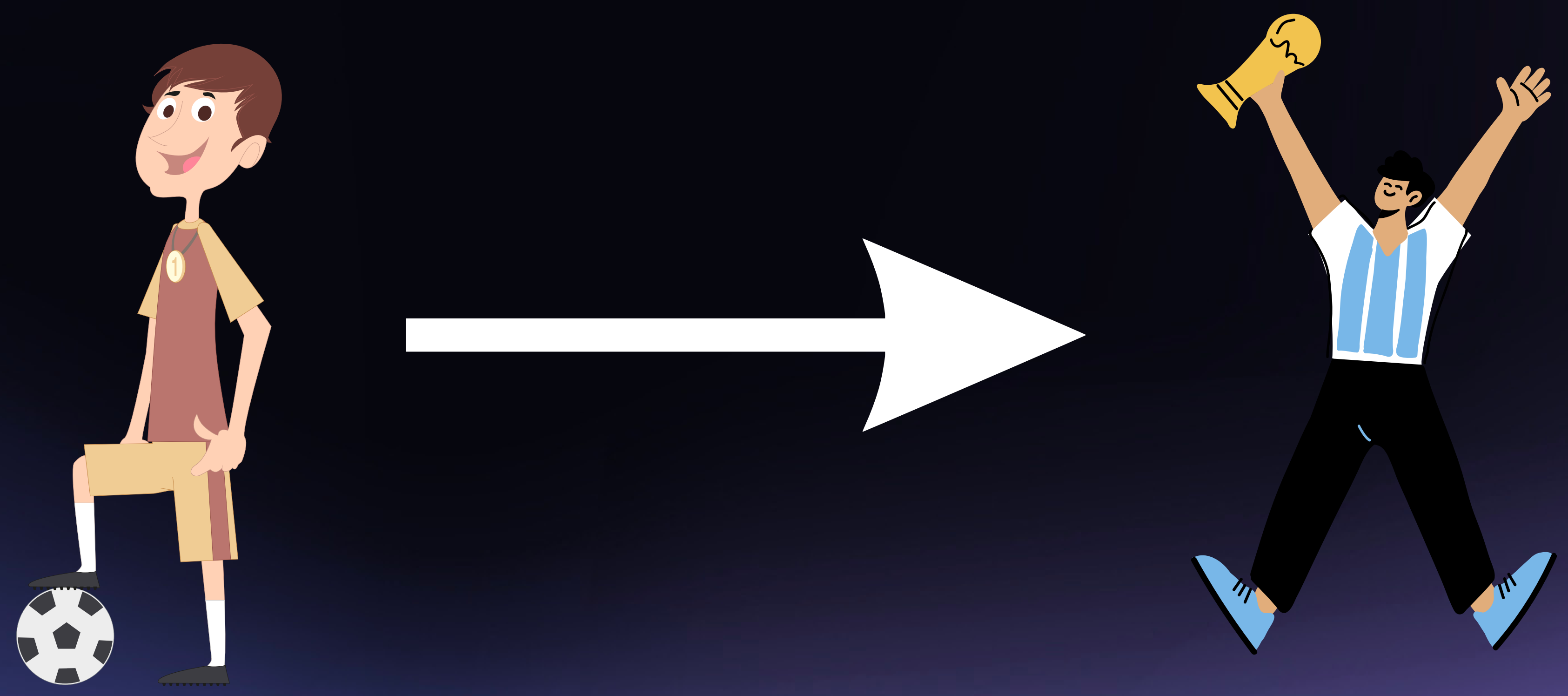
- Young soccer players need an average of 2.6s with a hit rate of 66% for the motor task in the Footbonaut.
- Players from the higher leagues who were playing junior performance centers achieve significantly lower action times with similar hit rates than players from the lower leagues.



Stein, Manuel, Janetzko, Halldor, Lamprecht, Andreas, Breitkreutz, Thorsten, Zimmermann, Philipp, Goldlucke, Bastian, Schreck, Tobias (Sport Analysis, 2018): Bring it to the Pitch: Combining Video and Movement Data to Enhance Team Sport Analysis, in: IEEE TVCG, 24 (2018), no. 1, p. 13-22.

Relevant outcomes for our project

- The training with the Footbonaut provides measurement results that can be used to distinguish higher-class players from amateur players.
- Democratizing the Footbonaut with our Mixed Reality application might help amateur football players improve their action times and hit rates.



Research objectives

SMALL-SIDED GAMES IN SOCCER: AMATEUR VS. PROFESSIONAL PLAYERS' PHYSIOLOGICAL RESPONSES, PHYSICAL, AND TECHNICAL ACTIVITIES

ALEXANDRE DELLAL,[1,2,3] STEPHEN HILL-HAAS,[4] CARLOS LAGO-PENAS,[5] AND KARIM CHAMARI[2]

- investigating the impact of small-sided games (SSGs) on professional versus amateur players in soccer regarding their technical capacities.
- examining how variations in SSGs, such as the number of allowed ball touches per possession, influence physiological responses, physical and technical load, and overall performance.

Dellal A, Hill-Haas S, Lago-Penas C, Chamari K. Small-sided games in soccer: amateur vs. professional players' physiological responses, physical, and technical activities. J Strength Cond Res. 2011 Sep;25(9):2371-81. doi: 10.1519/JSC.0b013e3181fb4296. PMID: 21869625.

Experimental design

- Elite and amateur soccer players participated in three different SSG formats (2 vs. 2, 3 vs. 3, 4 vs. 4) with varied number of ball touches (1 touch, 2 touches, or free play). SSG sessions were performed at the same time of the day with at least 2 days separating each session.
- The study included 20 international, elite soccer players and 20 amateur soccer players from the fourth French division.
- During the SSGs, each player's heart rate was continuously recorded. Blood lactate samples were collected at the third minute post-exercise. The rating of perceived exertion (RPE) was assessed using the scale proposed by Foster et al. Video recordings monitored successful passes, lost balls per possession, and total number of possessions. GPS technology tracked distances covered and speed of movement.

Dellal A, Hill-Haas S, Lago-Penas C, Chamari K. Small-sided games in soccer: amateur vs. professional players' physiological responses, physical, and technical activities. J Strength Cond Res. 2011 Sep;25(9):2371-81. doi: 10.1519/JSC.0b013e3181fb4296. PMID: 21869625.

Results

Amateurs completed fewer successful passes, recorded higher values for Rating of Perceived Exertion, lost more possessions, and covered less total distance in sprinting and high-intensity

2

Heart rates were similar between both. Amateurs executed more unsuccessful passes, lost more balls per possession, and were involved in fewer duels than professionals.

3

The time-motion activities for amateurs were significantly lower in comparison with professionals.

4

Very similar results to those observed in 3 vs. 3 SSG

Dellal A, Hill-Haas S, Lago-Penas C, Chamari K. Small-sided games in soccer: amateur vs. professional players' physiological responses, physical, and technical activities. J Strength Cond Res. 2011 Sep;25(9):2371-81. doi: 10.1519/JSC.0b013e3181fb4296. PMID: 21869625.

Relevant outcomes for our project

- Amateurs executed more unsuccessful passes, lost more balls per possession
- Hence, we can assume that professional football players perform more exercises, that train a fast action time, high accuracy of passing, and distinctive technique (such as training with the Footbonaut).
- By democratizing the Footbonaut, the performance gap between amateur and professional football players can be reduced.

Research objectives

Relationship between ball possession and match outcome in UEFA Champions League

Vinicius Martins Farias[1*], Wesley Bierhals Fernandes[2], Gabriel Gustavo Bergmann[2], Eraldo dos Santos Pinheiro[2]

- solving the problem of inconsistency and lack of clarity in understanding the relationship between ball possession and match outcomes in soccer, as different studies have found conflicting correlation patterns between ball possession and success in soccer.
- exploring this relationship in the context of the UEFA Champions League.

Farias, Vinícius & Fernandes, Wesley & Bergmann, Gabriel & Pinheiro, Eraldo. (2021). Relationship between ball possession and match outcome in UEFA Champions League. Motricidade. 16. 1-7. 10.6063/motricidade.18382.

Experimental design

- The study analyzed a total of 625 matches from seasons 2014/2015, 2015/2016, 2016/2017, 2017/2018, and 2018/2019 of the UEFA Champions League. Data was collected from the competition's official website and included statistics on ball possession and match results.
- The collected data was categorized based on the difference in ball possession percentages between the two teams in each match.
- Descriptive analysis was used to quantify matches won, drawn, and lost by teams with the most ball possession.

Farias, Vinícius & Fernandes, Wesley & Bergmann, Gabriel & Pinheiro, Eraldo. (2021). Relationship between ball possession and match outcome in UEFA Champions League. Motricidade. 16. 1-7. 10.6063/motricidade.18382.

Results

- The results of this research revealed a significant relationship between ball possession and match outcomes.
- These results indicate a trend where teams with a higher percentage of ball possession are more likely to win the match, especially in cases where ball possession was greater than 60%. This finding is consistent across five seasons of the UEFA Champions League.

Farias, Vinícius & Fernandes, Wesley & Bergmann, Gabriel & Pinheiro, Eraldo. (2021). Relationship between ball possession and match outcome in UEFA Champions League. Motricidade. 16. 1-7. 10.6063/motricidade.18382.

Relevant outcomes for our project

- Teams with higher ball possession generally win more matches overall in professional competitions.
- Fast action time, high accuracy of passing, and distinctive technique are crucial for high ball possession rates. All of these skills are trained in the Footbonaut.
- Improving these skills does not only lead to individual performance improvement, but also to better results to the overall team performance
- Democratizing the Footbonaut could also be helpful to trainers of amateur football clubs.

Agenda

01

Professional vs. Amateur Soccer Players

Physical, tactical and technical differences

02

Computer Vision in Sports

Extending human vision

03

XR in Sports

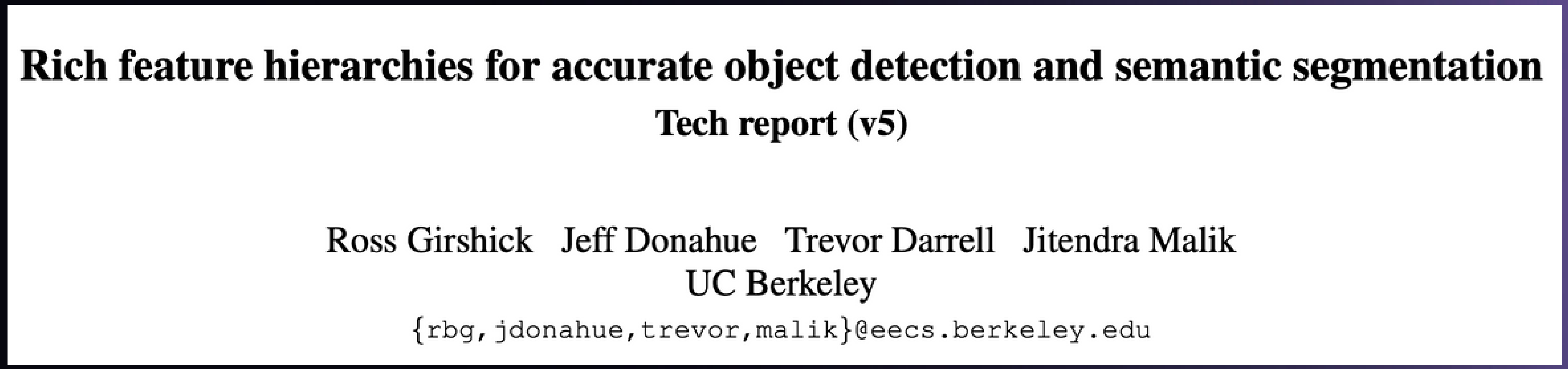
Merging virtual worlds with reality

04

Creating a MR Footbonaut

Democratizing the Footbonaut

Research objectives


Rich feature hierarchies for accurate object detection and semantic segmentation
Tech report (v5)

Ross Girshick Jeff Donahue Trevor Darrell Jitendra Malik
UC Berkeley
{rbg,jdonahue,trevor,malik}@eecs.berkeley.edu

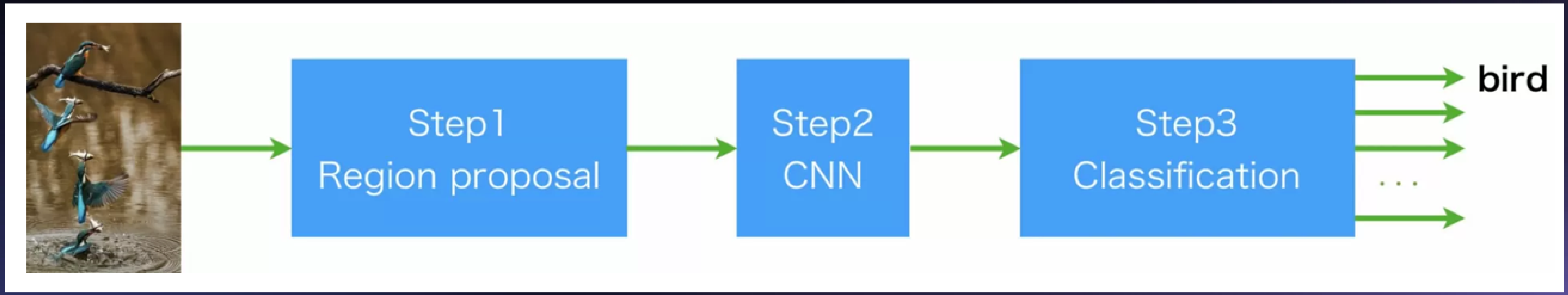
- The best performing methods were complex ensemble systems that combined multiple low-level image features with high-level context.
- To provide a simple, scalable detection algorithm that improves the average mean accuracy (mAP).
- Compute features that are more informative for visual recognition through a hierarchical, multi-stage process.

Ross Girshick, Jeff Donahue, Trevor Darrell, Jitendra Malik. Rich feature hierarchies for accurate object detection and semantic segmentation. 2014, arXiv:1311.2524

Method

3 main steps

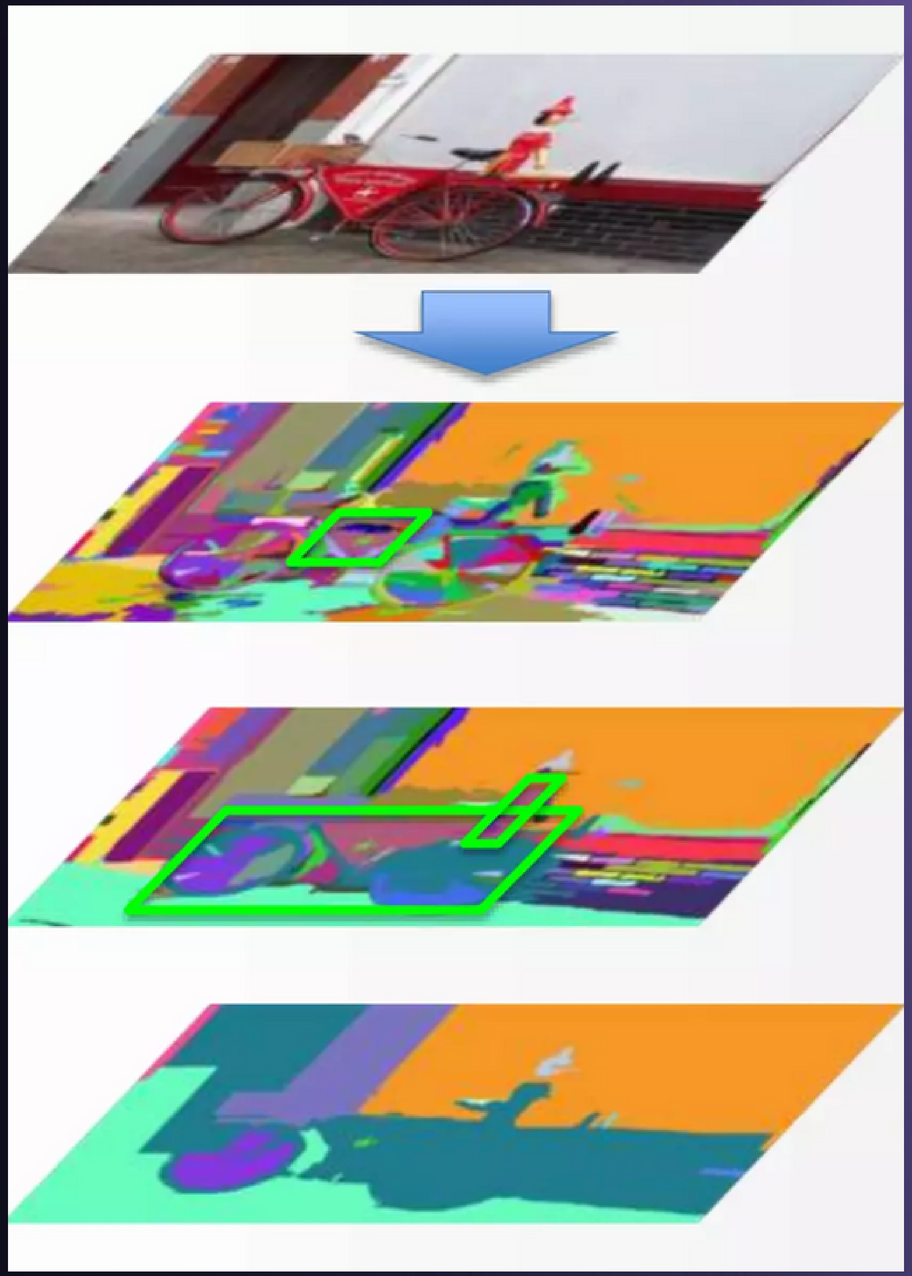
- Select the regions where the detectable object is expected to be captured (Selective Search)
- Extract features from selected regions using CNN
- Classify extracted features using a set of unique linear SVMs trained for each class

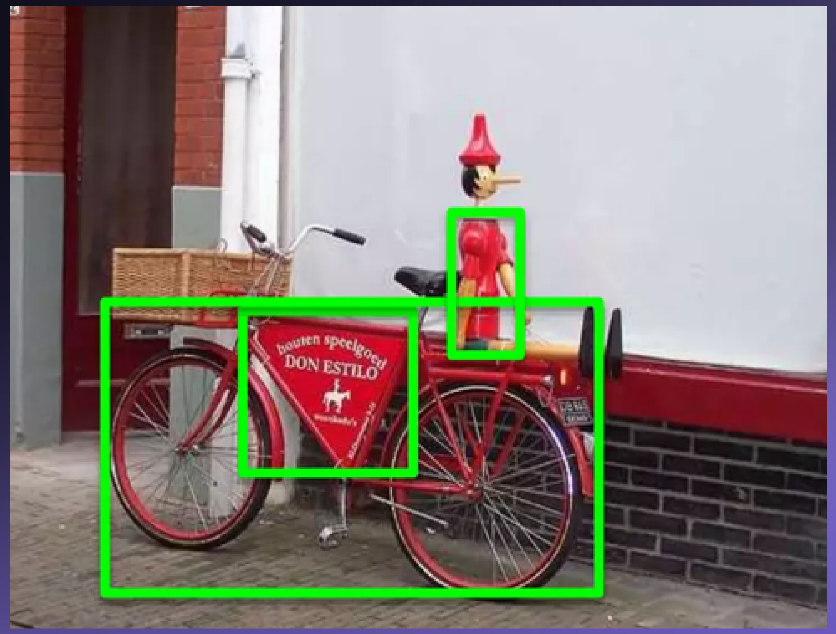


Method

Selective Search

- NOT learnable
- Segment similar regions at pixel level by Felzenszwalb algorithm
- Color histogram-based features are taken as a measure for calculating similarity of each region. (since each regions have different size, it's not possible to compare by pixel-based features)
- Combine regions with the neighbor which has the highest similarity
- Hierarchically run these steps until all pixels converge
- Take candidate areas from this process





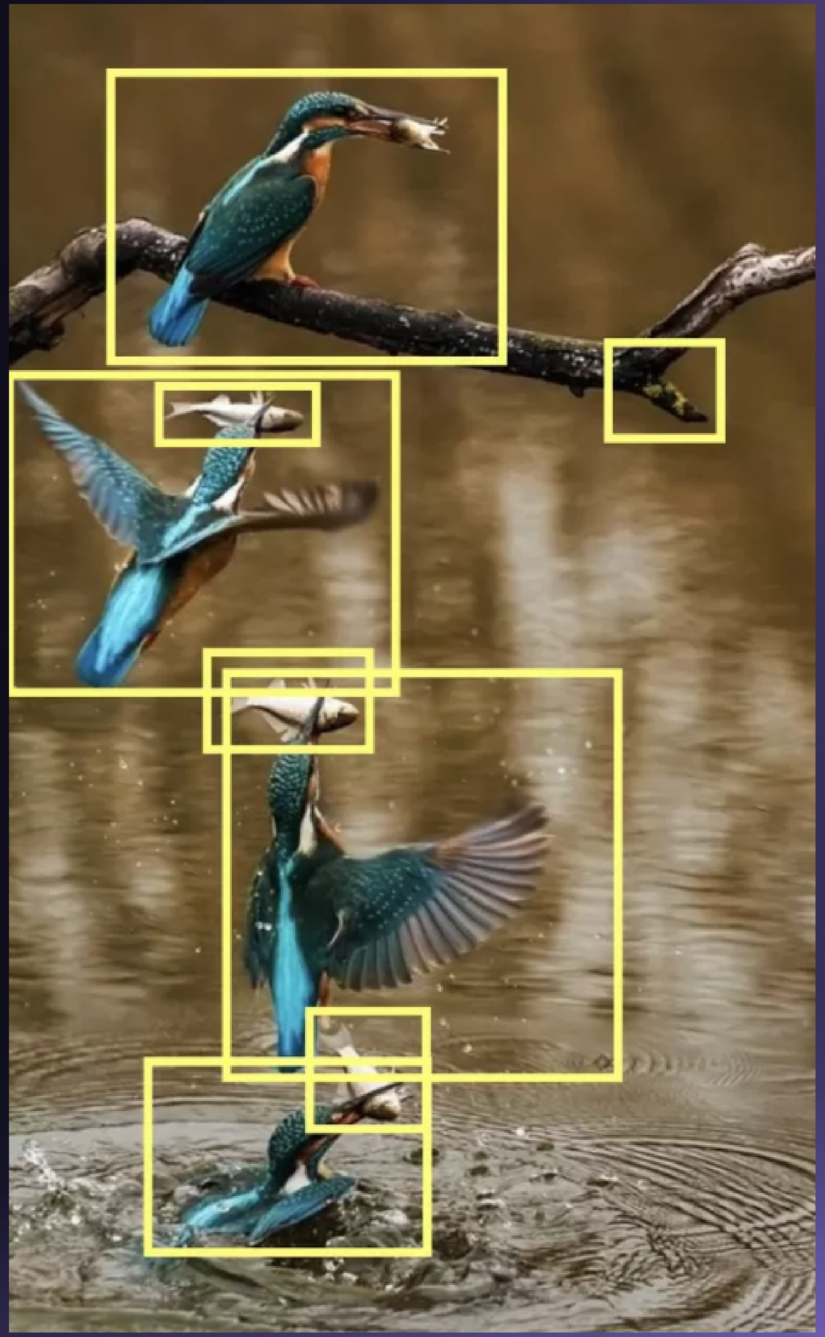
Method

Feature extraction

- Extract the feature of each rectangle area (output of Selective Search)
- Input partial images to CNN respectively

Classification

- Classify features with SVMs
- Each SVM only returns whether the detection target is contained or not
- The same number of SVMs must be prepared as classifiers as the number of detection objects



Results

- Evaluate with PASCAL VOC 2010/11, Compared to their multi-feature, non-linear kernel SVM approach, achieved a large improvement in mAP, from 35.1% to 53.7%.
- This method also achieved similar performance (53.3% mAP) on VOC 2011/12 test.

VOC 2010 test	aero	bike	bird	boat	bottle	bus	car	cat	chair	cow	table	dog	horse	mbike	person	plant	sheep	sofa	train	tv	mAP
DPM v5 [20]†	49.2	53.8	13.1	15.3	35.5	53.4	49.7	27.0	17.2	28.8	14.7	17.8	46.4	51.2	47.7	10.8	34.2	20.7	43.8	38.3	33.4
UVA [39]	56.2	42.4	15.3	12.6	21.8	49.3	36.8	46.1	12.9	32.1	30.0	36.5	43.5	52.9	32.9	15.3	41.1	31.8	47.0	44.8	35.1
Regionlets [41]	65.0	48.9	25.9	24.6	24.5	56.1	54.5	51.2	17.0	28.9	30.2	35.8	40.2	55.7	43.5	14.3	43.9	32.6	54.0	45.9	39.7
SegDPM [18]†	61.4	53.4	25.6	25.2	35.5	51.7	50.6	50.8	19.3	33.8	26.8	40.4	48.3	54.4	47.1	14.8	38.7	35.0	52.8	43.1	40.4
R-CNN	67.1	64.1	46.7	32.0	30.5	56.4	57.2	65.9	27.0	47.3	40.9	66.6	57.8	65.9	53.6	26.7	56.5	38.1	52.8	50.2	50.2
R-CNN BB	**71.8**	**65.8**	**53.0**	**36.8**	**35.9**	**59.7**	**60.0**	**69.9**	**27.9**	**50.6**	**41.4**	**70.0**	**62.0**	**69.0**	**58.1**	**29.5**	**59.4**	**39.3**	**61.2**	**52.4**	**53.7**

Table 1: Detection average precision (%) on VOC 2010 test. R-CNN is most directly comparable to UVA and Regionlets since all methods use selective search region proposals. Bounding-box regression (BB) is described in Section C. At publication time, SegDPM was the top-performer on the PASCAL VOC leaderboard. †DPM and SegDPM use context rescoring not used by the other methods.

Results

- Evaluate with ILSVRC2013, achieved a mAP of 31.4%, which is significantly ahead of the second-best result of 24.3% from OverFeat

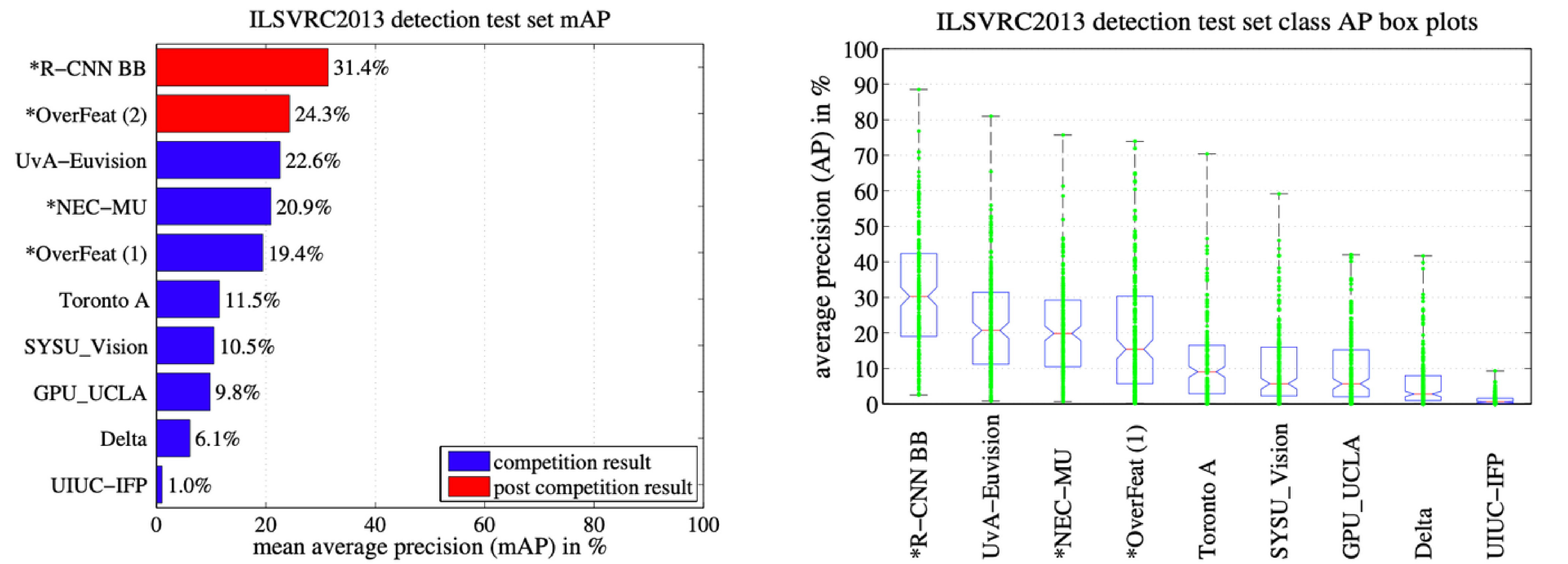


Figure 3: (Left) Mean average precision on the ILSVRC2013 detection test set. Methods preceeded by * use outside training data (images and labels from the ILSVRC classification dataset in all cases). **(Right) Box plots for the 200 average precision values per method.** A box plot for the post-competition OverFeat result is not shown because per-class APs are not yet available (per-class APs for R-CNN are in Table 8 and also included in the tech report source uploaded to arXiv.org; see `R-CNN-ILSVRC2013-APs.txt`). The red line marks the median AP, the box bottom and top are the 25th and 75th percentiles. The whiskers extend to the min and max AP of each method. Each AP is plotted as a green dot over the whiskers (best viewed digitally with zoom).

Discussion

- Most of the competing submissions (OverFeat, NEC-MU, UvA- Euvision, Toronto A, and UIUC-IFP) used CNNs, indicating that there is considerble nuance in how CNNs can be applied to object detection, leading to greatly varying outcomes.
- They have found that the choice of architecture has a large effect on R-CNN detection performance. R-CNN with OxfordNet substantially outperforms R-CNN with TorontoNet, increasing mAP from 58.5% to 66.0%. However there is a considerable drawback in terms of compute time, with the forward pass of O-Net taking roughly 7 times longer than T-Net.

Extension

Fast R-CNN

- To enable the model to be trained in a single training step, all structures after the CNN are neural networks and use multi-task loss.
- Generate fixed size feature map using RoI (Region of Interest) Pooling.
- Fast R-CNN is 9 times faster in training and 213 times faster in testing than R-CNN.
- Higher mAP than R-CNN.

Research objectives

You Only Look Once:
Unified, Real-Time Object Detection

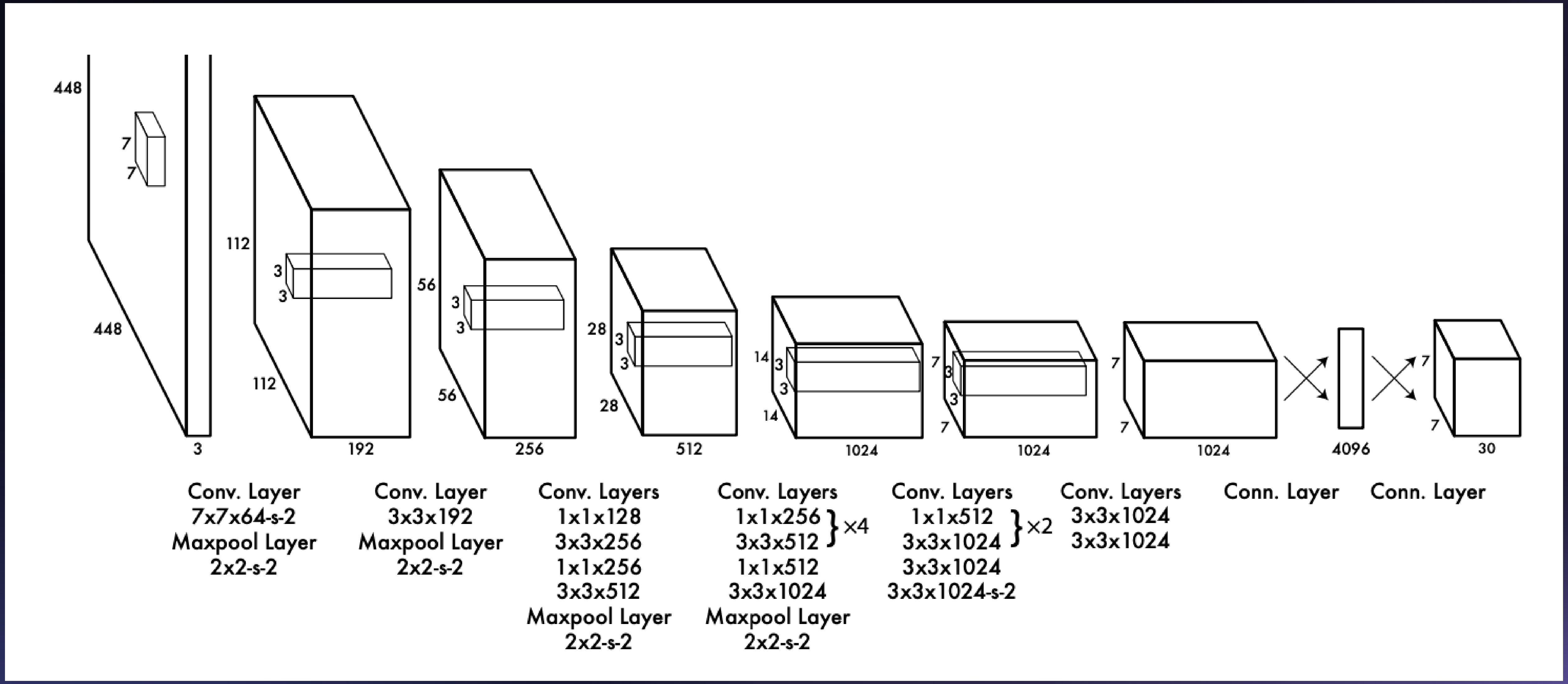
Joseph Redmon[*], Santosh Divvala[*†], Ross Girshick[¶], Ali Farhadi[*†]
University of Washington[*], Allen Institute for AI[†], Facebook AI Research[¶]

- Complex pipelines are slow and hard to optimize because each individual component must be trained separately.
- Reframe object detection as a single regression problem, straight from image pixels to bounding box coordinates and class probabilities. A single NN predicts it in one evaluation.
- To Enable end-to-end optimization directly related to detection performance

Joseph Redmon, Santosh Divvala, Ross Girshick, Ali Farhadi. You Only Look Once: Unified, Real-Time Object Detection. 2016, arXiv:1506.02640

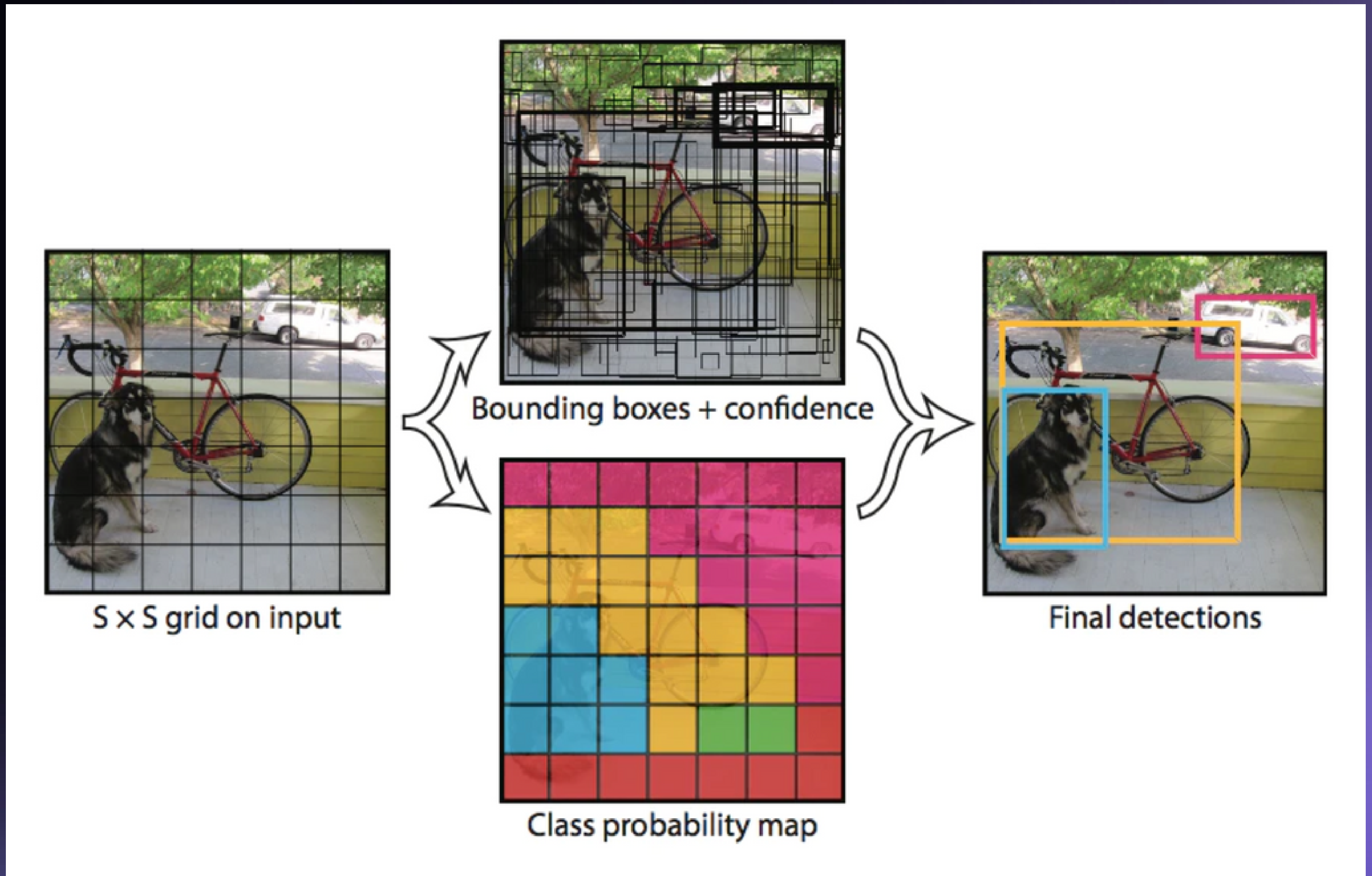
Method

- Region detection and classification are carried out in parallel
- The entire process is completed by a single CNN



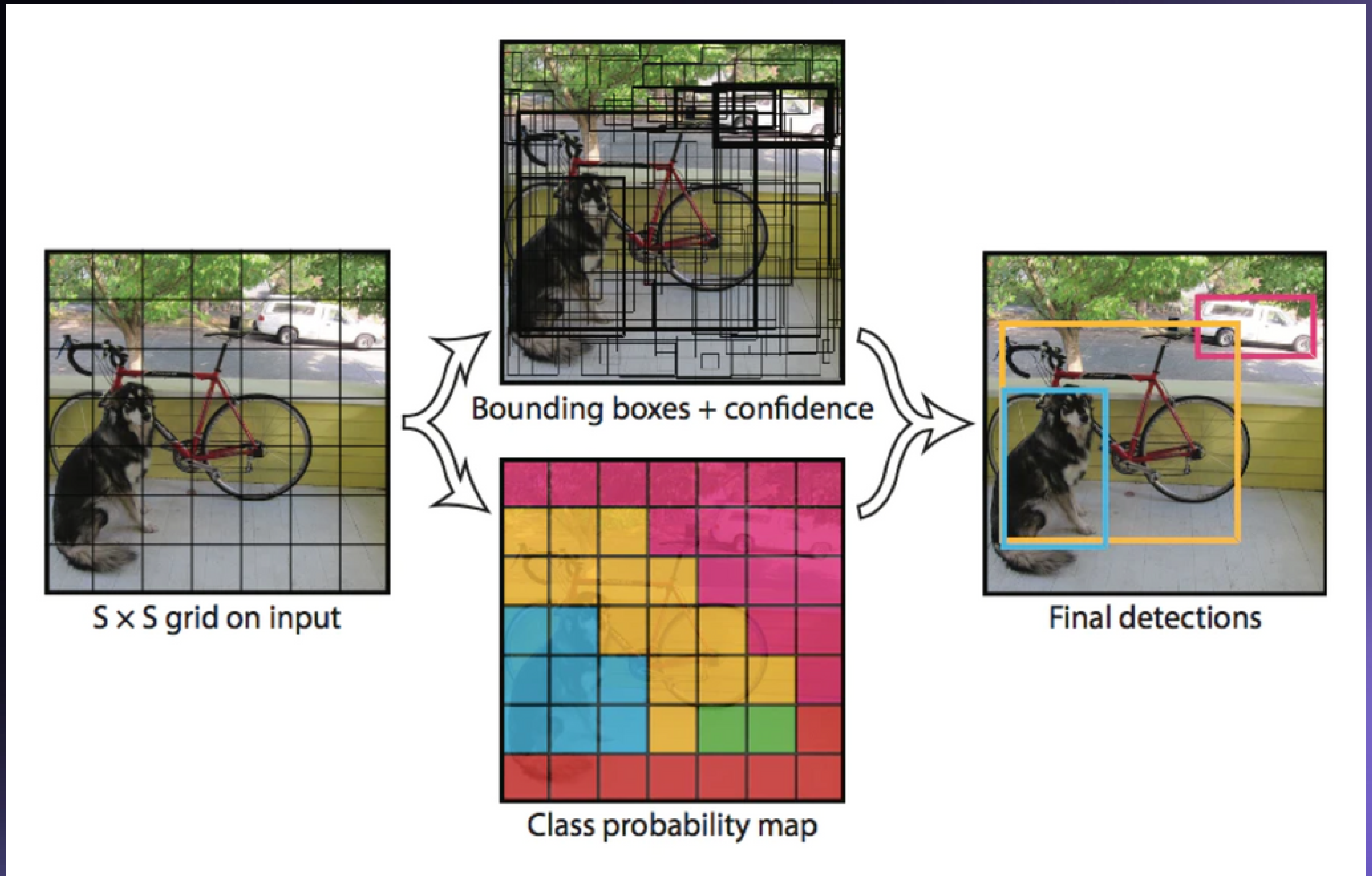
Method

- Divide an input image into SxS grids and define the maximum number of bounding boxes grid cell has is B
- Each Bounding box has its center within grid cell



Method

- Each bounding box output consisting of 4 corner positions and the confidence with which the detection target is contained
- Each grid cell output consisting classification probabilities of detection target
- Each provisional detections (SxSxB) are calculated by multiplying the bbox confidence and the classification probability as the output of network
- Obtain detections with confidence higher than the threshold as the final output



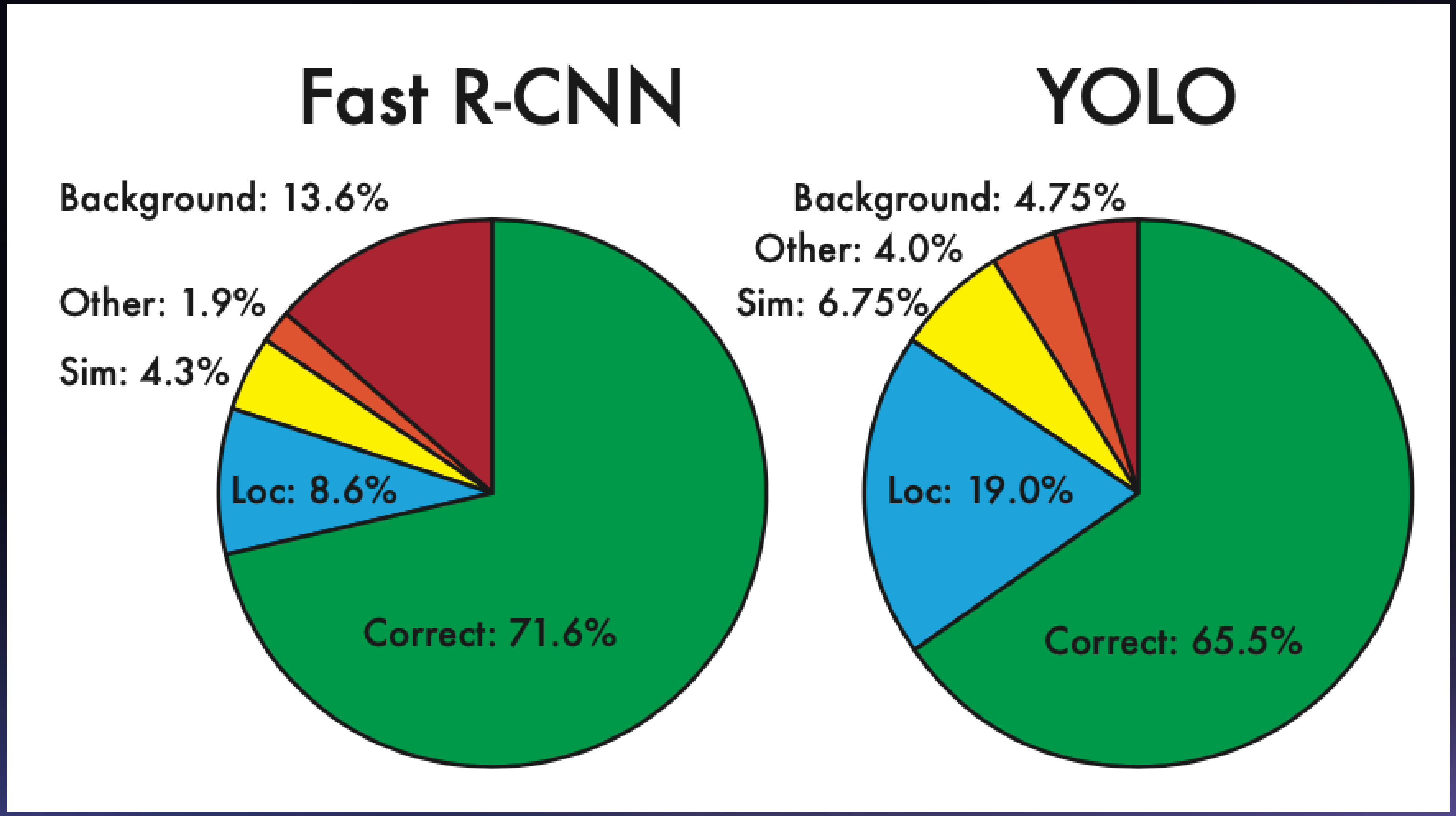
Method

- X,Y location error of bbox when any object is detected in the bbox
- W,H size error of bbox when any object is detected in the bbox
- 2 bbox confidence errors: when any object is detected and when none is detected (has its own coefficient)
- Classification error when any object is detected

$$\lambda_{\textbf{coord}} \sum_{i=0}^{S^2} \sum_{j=0}^{B} \mathbb{1}_{ij}^{\text{obj}} \left[(x_i - \hat{x}_i)^2 + (y_i - \hat{y}_i)^2 \right]$$
$$+ \lambda_{\textbf{coord}} \sum_{i=0}^{S^2} \sum_{j=0}^{B} \mathbb{1}_{ij}^{\text{obj}} \left[\left(\sqrt{w_i} - \sqrt{\hat{w}_i} \right)^2 + \left(\sqrt{h_i} - \sqrt{\hat{h}_i} \right)^2 \right]$$
$$+ \sum_{i=0}^{S^2} \sum_{j=0}^{B} \mathbb{1}_{ij}^{\text{obj}} \left(C_i - \hat{C}_i \right)^2$$
$$+ \lambda_{\text{noobj}} \sum_{i=0}^{S^2} \sum_{j=0}^{B} \mathbb{1}_{ij}^{\text{noobj}} \left(C_i - \hat{C}_i \right)^2$$
$$+ \sum_{i=0}^{S^2} \mathbb{1}_{i}^{\text{obj}} \sum_{c \in \text{classes}} (p_i(c) - \hat{p}_i(c))^2$$

Results

- Compare to SOTA detection systems, YOLO makes more localization errors but is less likely to predict false positives on background
- Since YOLO sees the entire image during training and test time so it implicitly encodes contextual information about classes as well as their appearance.



Results

- Achieve 88% on the ImageNet 2012 validation set
- Scores 57.9% mAP on the VOC 2012 validation set. This is lower than the current SOTA, closer to the original R-CNN using VGG-16
- Get a significant boost in performance by using YOLO to eliminate background detections from Fast R-CNN

VOC 2012 test	mAP	aero	bike	bird	boat	bottle	bus	car	cat	chair	cow	table	dog	horse	mbike	person	plant	sheep	sofa	train	tv
MR_CNN_MORE_DATA [11]	**73.9**	**85.5**	**82.9**	**76.6**	**57.8**	**62.7**	**79.4**	77.2	86.6	**55.0**	**79.1**	**62.2**	87.0	**83.4**	**84.7**	78.9	45.3	73.4	65.8	80.3	74.0
HyperNet_VGG	71.4	84.2	78.5	73.6	55.6	53.7	78.7	**79.8**	87.7	49.6	74.9	52.1	86.0	81.7	83.3	**81.8**	**48.6**	**73.5**	59.4	79.9	65.7
HyperNet_SP	71.3	84.1	78.3	73.3	55.5	53.6	78.6	79.6	87.5	49.5	74.9	52.1	85.6	81.6	83.2	81.6	48.4	73.2	59.3	79.7	65.6
Fast R-CNN + YOLO	70.7	83.4	78.5	73.5	55.8	43.4	79.1	73.1	**89.4**	49.4	75.5	57.0	**87.5**	80.9	81.0	74.7	41.8	71.5	68.5	**82.1**	67.2
MR_CNN_S_CNN [11]	70.7	85.0	79.6	71.5	55.3	57.7	76.0	73.9	84.6	50.5	74.3	61.7	85.5	79.9	81.7	76.4	41.0	69.0	61.2	77.7	72.1
Faster R-CNN [28]	70.4	84.9	79.8	74.3	53.9	49.8	77.5	75.9	88.5	45.6	77.1	55.3	86.9	81.7	80.9	79.6	40.1	72.6	60.9	81.2	61.5
DEEP_ENS_COCO	70.1	84.0	79.4	71.6	51.9	51.1	74.1	72.1	88.6	48.3	73.4	57.8	86.1	80.0	80.7	70.4	46.6	69.6	**68.8**	75.9	71.4
NoC [29]	68.8	82.8	79.0	71.6	52.3	53.7	74.1	69.0	84.9	46.9	74.3	53.1	85.0	81.3	79.5	72.2	38.9	72.4	59.5	76.7	68.1
Fast R-CNN [14]	68.4	82.3	78.4	70.8	52.3	38.7	77.8	71.6	89.3	44.2	73.0	55.0	**87.5**	80.5	80.8	72.0	35.1	68.3	65.7	80.4	64.2
UMICH_FGS_STRUCT	66.4	82.9	76.1	64.1	44.6	49.4	70.3	71.2	84.6	42.7	68.6	55.8	82.7	77.1	79.9	68.7	41.4	69.0	60.0	72.0	66.2
NUS_NIN_C2000 [7]	63.8	80.2	73.8	61.9	43.7	43.0	70.3	67.6	80.7	41.9	69.7	51.7	78.2	75.2	76.9	65.1	38.6	68.3	58.0	68.7	63.3
BabyLearning [7]	63.2	78.0	74.2	61.3	45.7	42.7	68.2	66.8	80.2	40.6	70.0	49.8	79.0	74.5	77.9	64.0	35.3	67.9	55.7	68.7	62.6
NUS_NIN	62.4	77.9	73.1	62.6	39.5	43.3	69.1	66.4	78.9	39.1	68.1	50.0	77.2	71.3	76.1	64.7	38.4	66.9	56.2	66.9	62.7
R-CNN VGG BB [13]	62.4	79.6	72.7	61.9	41.2	41.9	65.9	66.4	84.6	38.5	67.2	46.7	82.0	74.8	76.0	65.2	35.6	65.4	54.2	67.4	60.3
R-CNN VGG [13]	59.2	76.8	70.9	56.6	37.5	36.9	62.9	63.6	81.1	35.7	64.3	43.9	80.4	71.6	74.0	60.0	30.8	63.4	52.0	63.5	58.7
YOLO	57.9	77.0	67.2	57.7	38.3	22.7	68.3	55.9	81.4	36.2	60.8	48.5	77.2	72.3	71.3	63.5	28.9	52.2	54.8	73.9	50.8
Feature Edit [33]	56.3	74.6	69.1	54.4	39.1	33.1	65.2	62.7	69.7	30.8	56.0	44.6	70.0	64.4	71.1	60.2	33.3	61.3	46.4	61.7	57.8
R-CNN BB [13]	53.3	71.8	65.8	52.0	34.1	32.6	59.6	60.0	69.8	27.6	52.0	41.7	69.6	61.3	68.3	57.8	29.6	57.8	40.9	59.3	54.1
SDS [16]	50.7	69.7	58.4	48.5	28.3	28.8	61.3	57.5	70.8	24.1	50.7	35.9	64.9	59.1	65.8	57.1	26.0	58.8	38.6	58.9	50.7
R-CNN [13]	49.6	68.1	63.8	46.1	29.4	27.9	56.6	57.0	65.9	26.5	48.7	39.5	66.2	57.3	65.4	53.2	26.2	54.5	38.1	50.6	51.6

Results

- Base YOLO network runs at 45 frames per second with no batch processing on a Titan X GPU and
- Fast version YOLO runs at more than 150 fps
- Fast enough to process streaming video in real-time with less than 25 ms of latency

Real-Time Detectors	Train	mAP	FPS
100Hz DPM [31]	2007	16.0	100
30Hz DPM [31]	2007	26.1	30
Fast YOLO	2007+2012	52.7	**155**
YOLO	2007+2012	**63.4**	45
Less Than Real-Time			
Fastest DPM [38]	2007	30.4	15
R-CNN Minus R [20]	2007	53.5	6
Fast R-CNN [14]	2007+2012	70.0	0.5
Faster R-CNN VGG-16[28]	2007+2012	73.2	7
Faster R-CNN ZF [28]	2007+2012	62.1	18
YOLO VGG-16	2007+2012	66.4	21

Discussion

- YOLO still lags behind SOTA detection systems in accuracy.
- While it can quickly identify objects in images it struggles to precisely localize some objects, especially small ones.
- In any image, many grid cells do not contain objects, and the confidence of these cells can approach zero, dominating the gradients from cells containing objects. This can cause the model to become unstable and learning to diverge in the early stages.
- YOLO imposes strong spatial constraints on predictions since each grid cell only predicts two boxes and can only have one class.
- This spatial constraint limits the number of nearby objects that our model can predict. It makes predicting small objects appear in same area harder.

Detection method for objects moving fast

difficulty

- Due to the relatively long shutter times of consumer cameras, images of fast-moving subjects tend to have afterimage and be blurred, which reduces the accuracy of image recognition.
- It is difficult to detect fast moving objects only from a single frame.

Solutions

- Frame differences
- Use more than 2 frames as input
- Can be used for video domain

Relevant outcomes for our project

- R-CNN and YOLO are good architectures we can use for the ball detection of our project.
- We have to consider the Trade-Off between model size and performance.
- As a football is a fast moving object, we have to find out whether YOLO will produce acceptable results for our specific use-case.
- In the end, we should test both architectures with different parameter adjustments and compare their performances.

Research objectives

TrackNet: A Deep Learning Network for Tracking High-speed and Tiny Objects in Sports Applications

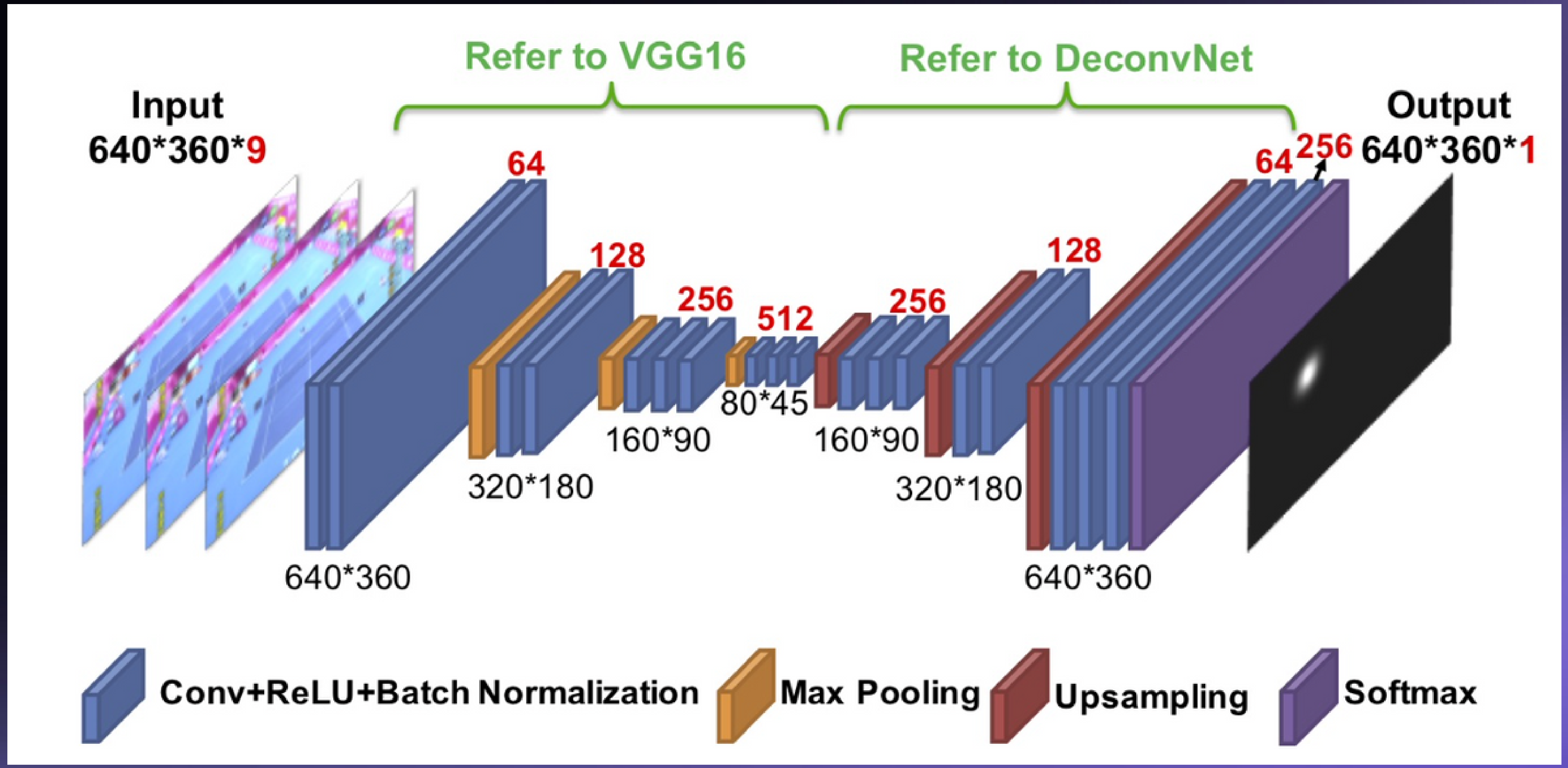
Yu-Chuan Huang I-No Liao Ching-Hsuan Chen Tsì-Uí İk* Wen-Chih Peng
Department of Computer Science, College of Computer Science
National Chiao Tung University
1001 University Road, Hsinchu City 30010, Taiwan
*Email: cwyi@nctu.edu.tw

- To track the tennis ball from broadcast videos in which the ball images are small, blurry, and sometimes with afterimage tracks or even invisible.
- To proposed heatmap-based deep learning method trained to not only recognize the ball image from a single frame but also learn flying patterns from consecutive frames.
- To unify the branched processes of previous methods.

Yu-Chuan Huang, I-No Liao, Ching-Hsuan Chen, Ts`ı-Uí˙Ik, Wen-Chih Peng. TrackNet: A Deep Learning Network for Tracking High-speed and Tiny Objects in Sports Applications. 2019, arXiv:1907.03698

Method

- Adopt the part of VGG-16 to generate the feature map
- Apply upsampling to generate heatmap
- The position of our target object is calculated based on the generated heatmap
- TrackNet is trained to generate a probability-like detection heatmap having the same resolution as the input frames
- Convert pixel-wisely the heatmap into a black-white binary heatmap with threshold (binarization)
- Exploit Hough Gradient Method to find the circle on binary-heatmap (if number of founded circle is 1, return centroid of it, otherwise, it's considered no ball detected)



Results

- In the tennis tracking application, base approach achieves outstanding performance of 95.7% precision, 89.6% recall, and 92.5% F1-measure
- By learning how to extract trajectory information from consecutive frames, TrackNet Model II further improves the performance and achieves 99.8% precision, 96.6% recall, and 98.2% F1-measure.

Model	Precision	Recall	F1-measure
Archana's [1]	92.5%	74.5%	82.5%
TrackNet Model I	95.7%	89.6%	92.5%
TrackNet Model II	99.8%	96.6%	98.2%
TrackNet Model II'	99.7%	97.3%	98.5%

Agenda

01

Professional vs. Amateur Soccer Players

Physical, tactical and technical differences

02

Computer Vision in Sports

Extending human vision

03

XR in Sports

Merging virtual worlds with reality

04

Creating a MR Footbonaut

Democratizing the Footbonaut

Research objectives

Research Article | Open Access
Volume 2022 | Article ID 1339434 | https://doi.org/10.1155/2022/1339434

Show citation

Analysis of the Application of Virtual Reality Technology in Football Training

Kun Zhao[1] and **Xueying Guo**[2]

- This paper focuses on researching various applications of VR technology in football training and analyzing the essential technologies that should be adopted in the VR training system.

Kun Zhao, Xueying Guo, "Analysis of the Application of Virtual Reality Technology in Football Training", Journal of Sensors, vol. 2022, Article ID 1339434, 8 pages, 2022. https://doi.org/10.1155/2022/1339434

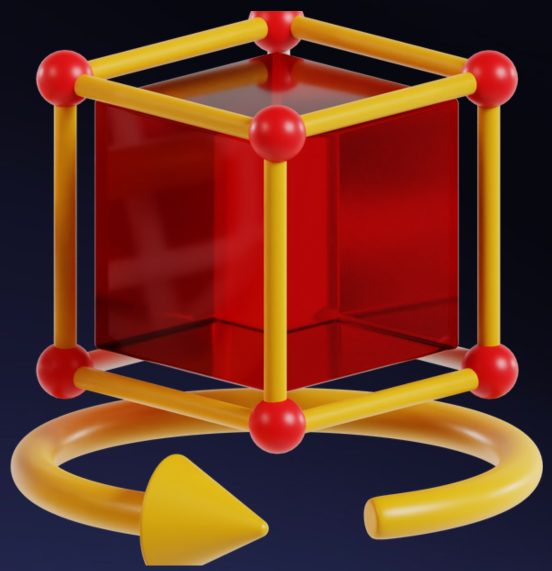
Experimental design

- Basic functions that a virtual football training system needs to have.
- Presentation of 2 VR Systems: the virtual simulated training ground and the desktop virtual football training system.
- Key technologies in the VR football training system
- Positive effects of the VR training system.

Kun Zhao, Xueying Guo, "Analysis of the Application of Virtual Reality Technology in Football Training", Journal of Sensors, vol. 2022, Article ID 1339434, 8 pages, 2022. https://doi.org/10.1155/2022/1339434

Basic Functions

realistic environment



collect athletes physiological data

free movement

post analysis

Kun Zhao, Xueying Guo, "Analysis of the Application of Virtual Reality Technology in Football Training", Journal of Sensors, vol. 2022, Article ID 1339434, 8 pages, 2022. https://doi.org/10.1155/2022/1339434

Virtual football training system & Desktop virtual football training system

- 3D scene design
- Character design
- Interactive controls

Kun Zhao, Xueying Guo, "Analysis of the Application of Virtual Reality Technology in Football Training", Journal of Sensors, vol. 2022, Article ID 1339434, 8 pages, 2022. https://doi.org/10.1155/2022/1339434

Key Technology in VR Football Training System

- Virtual Human Modelling
- Motion Data Information Capture (Player)
- Virtual Football Ground Design (Scale, Appearance)
- Interaction design (Intuitive and easy)

Kun Zhao, Xueying Guo, "Analysis of the Application of Virtual Reality Technology in Football Training", Journal of Sensors, vol. 2022, Article ID 1339434, 8 pages, 2022. https://doi.org/10.1155/2022/1339434

Results

- Improved Training: VR technology has significantly improved the training of football players, enhancing aspects like teamwork, better physical and tactical performance

- Psychological Benefits: VR technology's enhance the psychological readiness of players and improve their teamwork during matches.

Kun Zhao, Xueying Guo, "Analysis of the Application of Virtual Reality Technology in Football Training", Journal of Sensors, vol. 2022, Article ID 1339434, 8 pages, 2022. https://doi.org/10.1155/2022/1339434

Limitations

- Addressing VR and Reality Gap: VR simulations cannot yet capture all unexpected events in a real match, and adjustments are needed to better prepare players for such scenarios.

Kun Zhao, Xueying Guo, "Analysis of the Application of Virtual Reality Technology in Football Training", Journal of Sensors, vol. 2022, Article ID 1339434, 8 pages, 2022. https://doi.org/10.1155/2022/1339434

Relevant outcomes for our project

- Motion Data Information Capture, Virtual Football Ground Design (Scale, Appearance) and Interaction design are crucial for a good experience

- To address the gap between virtual training environments and the highly variable real-life playing conditions, Mixed Reality applications that merge real-world elements such as balls or team players with virtual objects such as marked shooting targets for a football can be used.

- Hence, developing a Mixed Reality Footbonaut not only has the potential to significantly improve the training routines of football players but also solves the inherent limitations of completely virtual training environments.

Research objectives

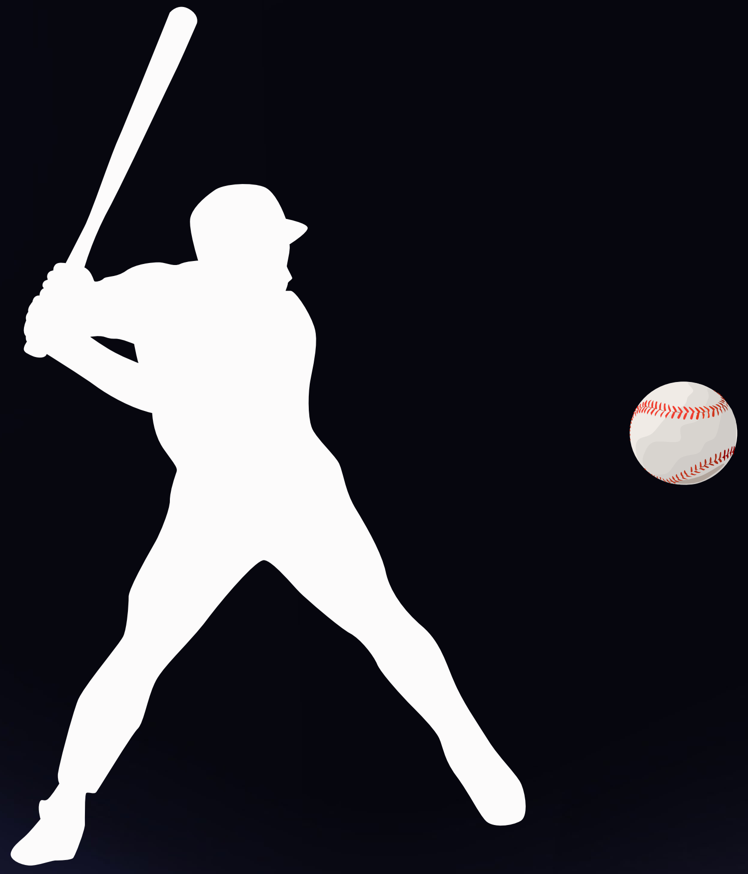
Front. Psychol., 13 December 2017
Sec. Movement Science and Sport Psychology
Volume 8 - 2017 | https://doi.org/10.3389/fpsyg.2017.02183

This article is part of the Research Topic
Current Issues in Perceptual Training: Facing the Requirement to Couple Perception, Cognition, and Action in Complex Motor Behavior
View all 14 Articles >

Transfer of Training from Virtual to Real Baseball Batting

Rob Gray*

- This study examined how VR Baseball training translates to real baseball performance.



Gray R. Transfer of Training from Virtual to Real Baseball Batting. Front Psychol. 2017 Dec 13;8:2183. doi: 10.3389/fpsyg.2017.02183. PMID: 29326627; PMCID: PMC5733365.

Experimental design

- Eighty baseball players were split into four groups, each undergoing a different form of training: adaptive hitting in the VE, extra sessions in the VE, extra real-life practice, or no additional training.
- The VE adaptive training used performance-based adjustments of pitch speed, type, and location.
- Pre- and post-training performance were evaluated through tests in the VE
- Additionally, the players' league batting statistics for the subsequent season and their highest competition level reached within five years post-training were analyzed.

Gray R. Transfer of Training from Virtual to Real Baseball Batting. Front Psychol. 2017 Dec 13;8:2183. doi: 10.3389/fpsyg.2017.02183. PMID: 29326627; PMCID: PMC5733365.

Results

The study found that VE adaptive training was superior in several ways:

- players showing significant improvement in 7 out of 8 batting performance assessments,
- superior on-base percentage (OBP) in the season following training
- a higher proportion reaching competition levels above high school compared to other groups.
- enhanced knowledge of the strike zone and superior pitch type recognition.

Gray R. Transfer of Training from Virtual to Real Baseball Batting. Front Psychol. 2017 Dec 13;8:2183. doi: 10.3389/fpsyg.2017.02183. PMID: 29326627; PMCID: PMC5733365.

Results

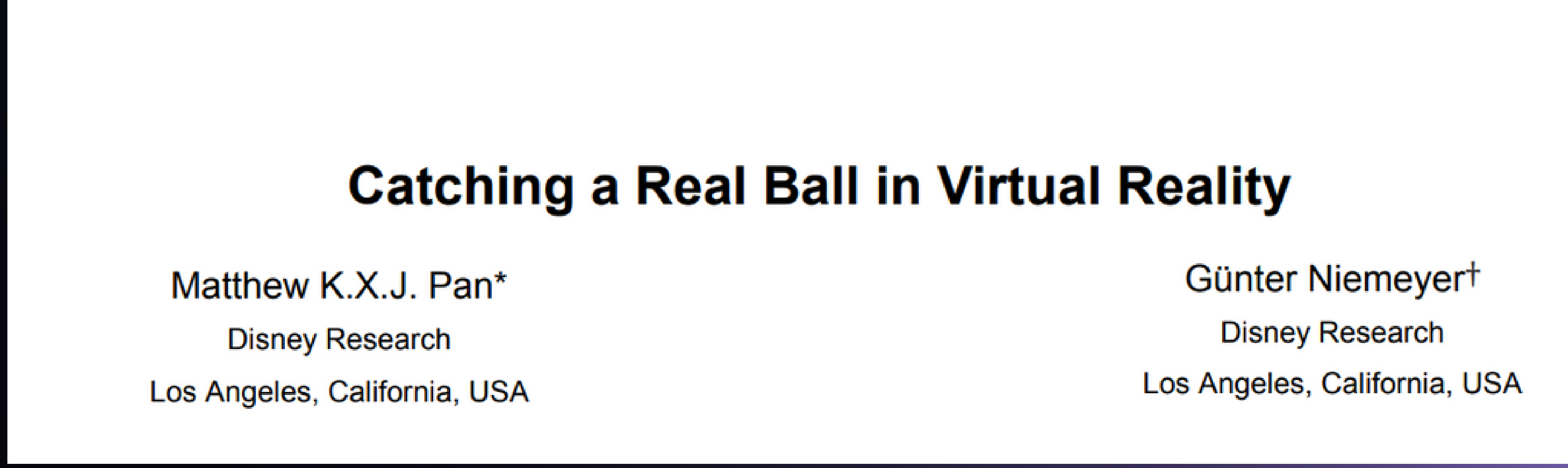
- significant improvements in both VE and real-life tests.
- This suggests that the VE adaptive training incorporated elements not usually present in real-life training, making it more effective.
- it was found that the adaptive training promoted greater sensitivity to visual information like lace rotation and direction of motion, helping players better predict pitch types and trajectories.

Gray R. Transfer of Training from Virtual to Real Baseball Batting. Front Psychol. 2017 Dec 13;8:2183. doi: 10.3389/fpsyg.2017.02183. PMID: 29326627; PMCID: PMC5733365.

Relevant outcomes for our project

- Besides the fact that XR experience could improve the performance of sports man over time, there might also occur other yet unnknown benefits e.g. dealing better with stress and anxiety.
- Factors like adaptive experience might have a high impact to the players progress also in football training.

Research objectives



Catching a Real Ball in Virtual Reality

Matthew K.X.J. Pan*
Disney Research
Los Angeles, California, USA

Günter Niemeyer†
Disney Research
Los Angeles, California, USA

Pan, M.K., & Niemeyer, G. (2017). Catching a real ball in virtual reality. 2017 IEEE Virtual Reality (VR), 269-270.

Experimental design

1. Virtual Ball: a virtual ball is rendered which tracks the real ball
2. Trajectory: the predicted trajectory of the real ball is displayed using a line
3. Target: a target-catching location lying on the predicted trajectory of the real ball is displayed.

To evaluate prediction accuracy, they tossed a ball using an underarm throw and continually estimated the horizontal target position and time to target with an Unscented Kalman Filter.

It takes the information from each video frame and makes an educated guess (estimate) about where the ball is, how fast it's moving, and in which direction.

Pan, M.K., & Niemeyer, G. (2017). Catching a real ball in virtual reality. 2017 IEEE Virtual Reality (VR), 269-270.

1. Initial state:

[x, y, z, vx, vy, vz]

2. Acceleration:

ax = -D*vx*sqrt(vx^2 + vy^2 + vz^2)

az = -g - D*vz*sqrt(vx^2 + vy^2 + vz^2)

acceleration due to gravity is only in the z direction, and the drag force is opposite to the direction of velocity

3. Update Velocities:

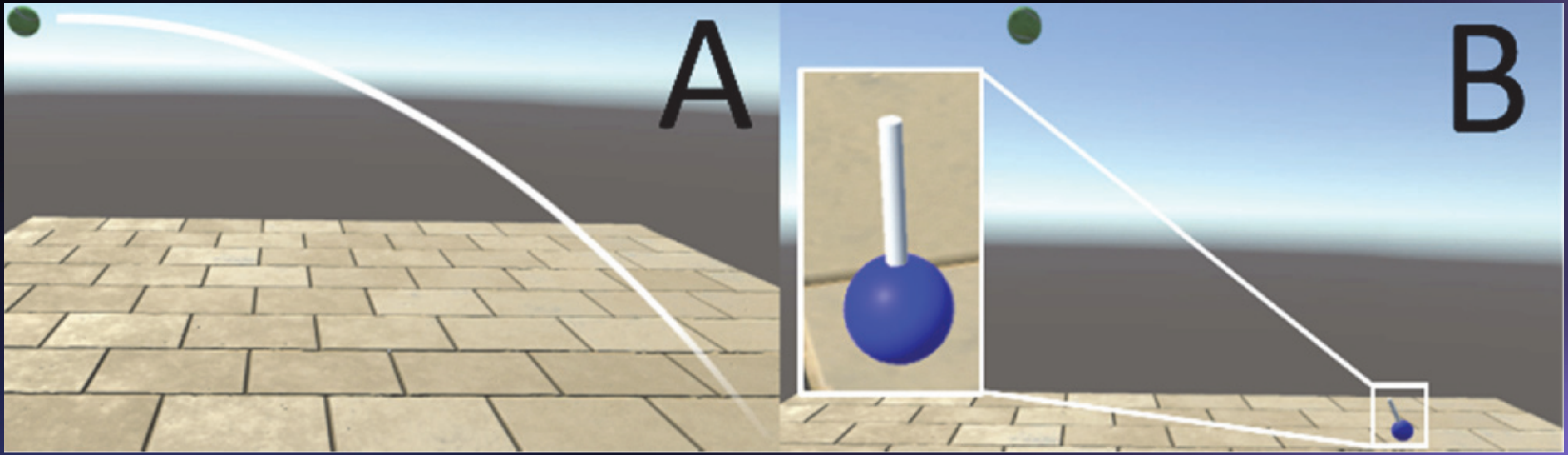
vx' = vx + ax*dt

4. Update Positions:

x' = x + vx*dt + 0.5*ax*dt^2

D = ρCA/2

where ρ is the density of air, C is the drag coefficient (approximated as 0.5 for a sphere) and A is the cross sectional area of the ball.



Pan, M.K., & Niemeyer, G. (2017). Catching a real ball in virtual reality. 2017 IEEE Virtual Reality (VR), 269-270.

Results

- The researchers tested the effectiveness of catching using their system over the course of 140 tosses
- In total, 132 balls were caught, underlining the overall success of the user’s ability to catch in VR.
- Tosses that were made with only the virtual ball which most closely matches how balls are caught in the physical world. In this condition, 95% of balls were caught, indicating that their system allows users to catch reliably.
- Their system allows users to be quite adept at catching balls while in VR. Thus, combining virtual and physical dynamic interactions to enrich virtual reality experiences is feasible.

Pan, M.K., & Niemeyer, G. (2017). Catching a real ball in virtual reality. 2017 IEEE Virtual Reality (VR), 269-270.

Relevant outcomes for our project

- It is possible to combine image processing techniques with Extended Reality to achieve a desired result.
- Real physical objects need to be detected and tracked with a high accuracy in order to use the position data for virtual experiences.

Pan, M.K., & Niemeyer, G. (2017). Catching a real ball in virtual reality. 2017 IEEE Virtual Reality (VR), 269-270.

Agenda

01

Professional vs. Amateur Soccer Players

Physical, tactical and technical differences

02

Computer Vision in Sports

Extending human vision

03

XR in Sports

Merging virtual worlds with reality

04

Creating a MR Footbonaut

Democratizing the Footbonaut

Our Approach