Summary Report: DeepSportradar-v1 Paper

https://arxiv.org/pdf/2208.08190v1

\*\*Title:\*\* DeepSportradar-v1: Computer Vision Dataset for Sports Understanding with High Quality Annotations

\*\*Authors:\*\* Gabriel Van Zandycke, Vladimir Somers, Maxime Istasse, Carlo Del Don, Davide Zambrano

\*\*Conference:\*\* MMSports ’22, ACM Multimedia 2022

# Objective

The paper introduces DeepSportradar-v1, a suite of high-resolution datasets and benchmarks aimed at advancing computer vision (CV) for sports analytics, specifically basketball. It provides raw sensor data, camera parameters, and annotations to bridge the gap between academic research and real-world applications.

# Datasets

- DeepSport dataset: Covers tasks such as 3D ball localization, camera calibration, and player segmentation.  
- DeepSportradar ReID dataset: Designed for player re-identification tasks.

# Tasks and Benchmarks

## 1. Ball 3D Localization

Goal: Estimate ball’s 3D position from a single calibrated image.  
Metric: Mean Absolute Diameter Error (MADE), Mean Absolute Projection Error (MAPE).  
Baseline: VGG16-based regression model.  
Result: MADE = 2.12 px, MAPE = 3.05m, MARE = 10%.

## 2. Camera Calibration

Goal: Predict camera projection parameters from a single image.  
Metric: Mean Squared Error (MSE) on 3D-2D point projections.  
Baseline: DeepLabv3-based segmentation + OpenCV calibration.  
Result: MSE ≈ 490–590 cm.

## 3. Player Instance Segmentation

Goal: Segment individual humans on or near the court.  
Metric: Mean Average Precision (mAP) based on segmentation IoU.  
Baseline: Mask R-CNN (ResNeXt-101).  
Result: mAP = 0.51.

## 4. Player Re-Identification

Goal: Re-identify players across frames captured from the same camera.  
Metric: mAP, Rank-1, and Rank-5 accuracy.  
Baseline: ResNet-50 with Open-ReID.  
Result: mAP = 65%, Rank-1 = 90%, Rank-5 = 96%.

# Key Contributions

- Public release of high-resolution basketball datasets with raw sensor data.  
- Defined four challenging CV tasks with clear benchmarks.  
- Encourages reproducible research through open-source toolkits.  
- Hosts public competition to drive progress in CV for sports.

# Significance

Unlike previous datasets like SoccerNet, DeepSportradar-v1 provides access to original camera parameters and high-resolution data, making it more suitable for real-world deployment in sports analytics.

**Reference Papers**

**[1] Md Zahangir Alom et al., 2019 – “A State-of-the-Art Survey on Deep Learning Theory and Architectures”**

[**https://arxiv.org/abs/1807.04226**](https://arxiv.org/abs/1807.04226)

**Key Concepts and Taxonomy:**  
This paper surveys deep learning (DL) architectures including CNNs, RNNs, autoencoders, GANs, and reinforcement learning. It also discusses theoretical advancements, training techniques, and optimization challenges.

**Main Contributions and Findings:**  
It categorizes modern DL architectures, providing detailed comparisons and insights into their design principles, advantages, and typical application domains.

**Limitations and Future Directions:**  
The survey does not deeply evaluate task-specific benchmarks or real-time deployment issues. It encourages further work on interpretability, energy-efficient models, and generalization across domains.

**Relevance to Primary Paper:**  
Provides foundational context for DL models used in the DeepSportradar benchmark tasks, particularly segmentation and ReID tasks which rely on advanced neural architectures.

**[2] Daniel Bolya et al., 2019 – “YOLACT: Real-Time Instance Segmentation”**

[**https://arxiv.org/abs/1807.04226**](https://arxiv.org/abs/1807.04226)

**https://github.com/dbolya/yolact**

**Key Concepts and Taxonomy:**  
YOLACT introduces a one-stage instance segmentation framework designed for real-time processing by decoupling mask generation from object detection.

**Main Contributions and Findings:**  
Achieves fast inference (~33 fps) with competitive accuracy using parallel prediction of prototype masks and per-instance mask coefficients.

**Limitations and Future Directions:**  
While fast, accuracy lags behind two-stage methods like Mask R-CNN. Future work could address the tradeoff between speed and segmentation precision.

**Relevance to Primary Paper:**  
The real-time performance of YOLACT is relevant to instance segmentation tasks in DeepSportradar where processing speed is critical in sports analytics.

**[3] Jianhui Chen and James J Little, 2019 – “Sports Camera Calibration via Synthetic Data”**

**https://github.com/dbolya/yolact**

**Key Concepts and Taxonomy:**  
This work explores camera calibration for sports fields using synthetic datasets, leveraging prior knowledge of the playing area layout.

**Main Contributions and Findings:**  
Demonstrates the feasibility of using synthetic courts to train models that generalize to real-world data for camera calibration.

**Limitations and Future Directions:**  
Generalization across different sports or camera viewpoints remains a challenge. The domain gap between synthetic and real images needs to be further minimized.

**Relevance to Primary Paper:**  
Supports the camera calibration task in DeepSportradar, aligning with its use of court models and known geometry for automatic calibration.

**[4] Kai Chen et al., 2019 – “MMDetection: Open MMLab Detection Toolbox and Benchmark”**

[**https://github.com/dbolya/yolact**](https://github.com/dbolya/yolact)

**https://github.com/dbolya/yolact**

**Key Concepts and Taxonomy:**  
MMDetection is a flexible and modular open-source platform for object detection and instance segmentation, supporting many state-of-the-art methods.

**Main Contributions and Findings:**  
Provides a unified framework with strong baselines and reproducible benchmarks across diverse architectures like Faster R-CNN, Mask R-CNN, RetinaNet.

**Limitations and Future Directions:**  
Complexity can be a barrier for beginners. More intuitive APIs and real-time deployment support are potential areas for development.

**Relevance to Primary Paper:**  
DeepSportradar’s segmentation baseline is built on MMDetection, benefiting directly from its flexibility and robust implementations.

**[5] Liang-Chieh Chen et al., 2017 – “Rethinking Atrous Convolution for Semantic Image Segmentation”**

**https://github.com/dbolya/yolact**

**Key Concepts and Taxonomy:**  
Introduces atrous (dilated) convolution and DeepLab variants, enhancing receptive fields without increasing parameter count for dense predictions.

**Main Contributions and Findings:**  
DeepLab models achieve strong segmentation results, particularly in preserving spatial resolution in high-level feature maps.

**Limitations and Future Directions:**  
Atrous convolution can lead to grid artifacts. Future architectures might combine it with attention mechanisms or adaptive kernels.

**Relevance to Primary Paper:**  
DeepSportradar uses DeepLabv3 for court line segmentation in camera calibration, leveraging its ability to segment fine-grained structures.

**[6] Bowen Cheng et al., 2020 – “Panoptic-DeepLab: A Simple, Strong, and Fast Baseline for Bottom-Up Panoptic Segmentation”**

[**https://github.com/dbolya/yolact**](https://github.com/dbolya/yolact)

**https://github.com/facebookresearch/detectron2/tree/main/projects/Panoptic-DeepLab**

**Key Concepts and Taxonomy:**  
Panoptic-DeepLab tackles the panoptic segmentation task by combining semantic segmentation and instance segmentation using a bottom-up approach.

**Main Contributions and Findings:**  
The model achieves state-of-the-art performance with efficient inference by predicting center points and offsets for each object instance.

**Limitations and Future Directions:**  
Performance in crowded scenes or where object boundaries are unclear remains limited. Future models may benefit from enhanced spatial reasoning or contextual aggregation.

**Relevance to Primary Paper:**  
Relevant to the player instance segmentation task in DeepSportradar, particularly in crowded basketball court scenes where player boundaries may overlap.

**[7] Anthony Cioppa et al., 2021 – “Camera Calibration and Player Localization in SoccerNet-V2 and Investigation of Their Representations for Action Spotting”**

**https://arxiv.org/abs/2105.04825**

**Key Concepts and Taxonomy:**  
Proposes camera calibration and player localization methods in broadcast soccer videos using SoccerNet-v2 dataset.

**Main Contributions and Findings:**  
Develops a framework to extract calibration parameters and localize players in 2D, improving downstream action spotting tasks.

**Limitations and Future Directions:**  
Limited to broadcast views and 2D player localization; does not cover 3D reconstruction or multi-camera fusion.

**Relevance to Primary Paper:**  
Serves as a foundational reference for DeepSportradar’s calibration task; contrasts DeepSportradar’s use of raw high-res images and precise sensor metadata.

**[8] Anthony Cioppa et al., 2022 – “Scaling Up SoccerNet with Multi-View Spatial Localization and Re-Identification”**

**https://openaccess.thecvf.com/content/CVPR2022W/CVSports/html/Cioppa\_Scaling\_Up\_SoccerNet\_With\_Multi-View\_Spatial\_Localization\_and\_Re-Identification\_CVPRW\_2022\_paper.html**

**Key Concepts and Taxonomy:**  
Expands SoccerNet to support multi-view localization and ReID, incorporating spatial and temporal consistency.

**Main Contributions and Findings:**  
Introduces a benchmark for player tracking and identification across camera views, enhancing soccer video understanding.

**Limitations and Future Directions:**  
Mostly focused on soccer; generalization to other sports like basketball is untested. Also relies on limited camera viewpoints.

**Relevance to Primary Paper:**  
DeepSportradar builds on similar goals but focuses on basketball with richer raw data, highlighting the need for sport-specific datasets.

**[9] Anthony Cioppa et al., 2022 – “SoccerNet-Tracking: Multiple Object Tracking Dataset and Benchmark in Soccer Videos”**

**https://arxiv.org/abs/2205.14389**

**Key Concepts and Taxonomy:**  
Proposes a large-scale MOT benchmark for soccer, combining object detection and tracking in broadcast footage.

**Main Contributions and Findings:**  
Provides a challenging benchmark for evaluating trackers on fast-moving, occluded players with consistent identities over time.

**Limitations and Future Directions:**  
Limited by camera motion and broadcast constraints. Scene-specific features like lighting or graphics overlays can interfere with tracking.

**Relevance to Primary Paper:**  
Highlights the limitations of broadcast-based datasets; DeepSportradar addresses these by using sensor-aligned, high-quality camera footage.

**[10] Marius Cordts et al., 2016 – “The Cityscapes Dataset for Semantic Urban Scene Understanding”**

[**https://arxiv.org/abs/1604.01685**](https://arxiv.org/abs/1604.01685)

**https://www.cityscapes-dataset.com/**

**Key Concepts and Taxonomy:**  
Cityscapes is a benchmark dataset for pixel-level semantic segmentation in urban street scenes, widely used in autonomous driving research.

**Main Contributions and Findings:**  
Introduces dense annotations, standardized metrics, and varied real-world driving scenarios for evaluating segmentation models.

**Limitations and Future Directions:**  
Focused solely on street scenes; generalization to other domains (like sports) is limited. Also, it does not address instance segmentation explicitly.

**Relevance to Primary Paper:**  
Provides benchmarking methodology and annotation strategies that influenced datasets like DeepSport for player instance segmentation.

**[11] Adrien Deliege et al., 2021 – “SoccerNet-v2: A Dataset and Benchmarks for Holistic Understanding of Broadcast Soccer Videos”**

**https://arxiv.org/abs/2110.03923**

**Key Concepts and Taxonomy:**  
SoccerNet-v2 expands upon SoccerNet with fine-grained annotations for action spotting, camera calibration, and player localization.

**Main Contributions and Findings:**  
Introduces a comprehensive dataset and benchmark suite for multiple CV tasks in soccer using broadcast video, enabling joint scene understanding.

**Limitations and Future Directions:**  
Broadcast-specific constraints (camera motion, occlusions, low resolution) limit applicability for fine-grained 3D tasks or fast player actions.

**Relevance to Primary Paper:**  
DeepSportradar addresses these limitations by providing high-quality raw images and sensor data for basketball, improving precision for tasks like calibration and ReID.

**[12] Dirk Farin et al., 2003 – “Robust Camera Calibration for Sport Videos Using Court Models”**

**https://ieeexplore.ieee.org/document/1266522**

**Key Concepts and Taxonomy:**  
Early method for calibrating sports video using geometric models of the playing field, enabling 2D-to-3D mapping without direct measurements.

**Main Contributions and Findings:**  
Demonstrates that knowledge of court lines and structure can be used to extract camera parameters reliably.

**Limitations and Future Directions:**  
Manual steps and reliance on visible lines limit automation. Not adaptable to scenes with partial occlusions or poor line visibility.

**Relevance to Primary Paper:**  
Lays the groundwork for automated calibration methods in DeepSportradar, which extends these ideas using deep learning and dense annotation.

**[13] Dengpan Fu et al., 2021 – “Unsupervised Pre-Training for Person Re-Identification”**

**https://arxiv.org/abs/2006.08349**

**Key Concepts and Taxonomy:**  
Explores self-supervised learning for person ReID, where models are pre-trained without labeled identity data using contrastive objectives.

**Main Contributions and Findings:**  
Shows that unsupervised pre-training can improve downstream ReID performance, particularly when labeled data is scarce.

**Limitations and Future Directions:**  
Still underperforms compared to supervised approaches on complex datasets; requires careful augmentation strategies.

**Relevance to Primary Paper:**  
Relevant to DeepSportradar’s ReID task, where visual similarity among players makes robust representation learning crucial—especially when data annotations are limited.

**[14] Yixiao Ge et al., 2020 – “Self-Paced Contrastive Learning with Hybrid Memory for Domain Adaptive Object Re-ID”**

**https://arxiv.org/abs/2004.14546**

**Key Concepts and Taxonomy:**  
Proposes a self-paced domain adaptation method using hybrid memory modules for ReID tasks across different domains.

**Main Contributions and Findings:**  
Enables learning discriminative features under domain shifts, improving model transferability with minimal supervision.

**Limitations and Future Directions:**  
Memory-based approaches introduce complexity and may require fine-tuning of hyperparameters across datasets.

**Relevance to Primary Paper:**  
Could enhance DeepSportradar’s ReID task by enabling adaptation to new arenas or lighting conditions without requiring extensive labeled data.

**[15] Silvio Giancola et al., 2018 – “SoccerNet: A Scalable Dataset for Action Spotting in Soccer Videos”**

[**https://arxiv.org/abs/2004.14546**](https://arxiv.org/abs/2004.14546)

**https://www.soccer-net.org/**

**Key Concepts and Taxonomy:**  
SoccerNet introduces large-scale soccer video annotations for spotting key actions (goals, cards, etc.) in broadcast footage.

**Main Contributions and Findings:**  
Provides weakly-supervised labels for temporal action detection in long videos, enabling event-based indexing and retrieval.

**Limitations and Future Directions:**  
Temporal annotations are sparse and may not capture fine-grained spatial relationships. Dataset is limited to soccer and broadcast media.

**Relevance to Primary Paper:**  
Demonstrates the importance of task-specific sports datasets; DeepSportradar generalizes this idea with denser, spatially-aligned basketball data for CV tasks.

**[16] Agrim Gupta et al., 2019 – “LVIS: A Dataset for Large Vocabulary Instance Segmentation”**

[**https://arxiv.org/abs/1908.03195**](https://arxiv.org/abs/1908.03195)

**https://www.lvisdataset.org/**

**Key Concepts and Taxonomy:**  
LVIS introduces a large-scale dataset for instance segmentation with a long-tail distribution across 1000+ object categories.

**Main Contributions and Findings:**  
Highlights the challenge of class imbalance in segmentation tasks and proposes metrics better suited for evaluating rare class performance.

**Limitations and Future Directions:**  
Focused on general object segmentation, not domain-specific cases like sports. Annotation complexity and sparsity in rare classes can hinder some methods.

**Relevance to Primary Paper:**  
Provides insights on instance segmentation strategies that inform DeepSportradar’s single-class segmentation task (players), especially regarding occlusion and pose diversity.

**[17] Kaiming He et al., 2017 – “Mask R-CNN”**

**https://arxiv.org/abs/1703.06870**

**Key Concepts and Taxonomy:**  
Mask R-CNN extends Faster R-CNN by adding a parallel branch for predicting object masks, enabling instance-level segmentation.

**Main Contributions and Findings:**  
Achieves state-of-the-art accuracy on instance segmentation benchmarks with flexible, modular design and end-to-end training.

**Limitations and Future Directions:**  
Relatively slow inference; struggles with fine-grained boundaries and crowded object regions without refinement modules.

**Relevance to Primary Paper:**  
Used as the baseline for player instance segmentation in DeepSportradar, offering strong performance and broad community support.

**[18] Kaiming He et al., 2016 – “Deep Residual Learning for Image Recognition”**

**https://arxiv.org/abs/1703.06870**

**Key Concepts and Taxonomy:**  
Introduces ResNet architecture using residual connections to ease the training of deep neural networks.

**Main Contributions and Findings:**  
Demonstrated that very deep models (up to 152 layers) could be trained effectively, leading to significant improvements in image classification and transfer learning.

**Limitations and Future Directions:**  
High resource demands; performance gains taper off beyond certain depths. Newer models like transformers have started to challenge its dominance.

**Relevance to Primary Paper:**  
ResNet-50 serves as the backbone for the ReID baseline in DeepSportradar, offering a strong starting point for feature embedding.

**[19] Lingxiao He et al., 2020 – “FastReID: A PyTorch Toolbox for General Instance Re-Identification”**

[**https://arxiv.org/abs/1703.06870**](https://arxiv.org/abs/1703.06870)

**https://github.com/JDAI-CV/fast-reid**

**Key Concepts and Taxonomy:**  
FastReID is a modular and high-performance PyTorch library designed to streamline training and benchmarking of ReID models.

**Main Contributions and Findings:**  
Supports a wide range of state-of-the-art methods and loss functions, making it easy to benchmark across datasets and architectures.

**Limitations and Future Directions:**  
Primarily supports image-based ReID; does not directly address video-based or long-term tracking tasks.

**Relevance to Primary Paper:**  
Inspires the baseline design for DeepSportradar’s ReID task, offering tools and ideas for extensible, reproducible experiments.

**[20] Namdar Homayounfar et al., 2017 – “Sports Field Localization via Deep Structured Models”**

**https://openaccess.thecvf.com/content\_cvpr\_2017/html/Homayounfar\_Sports\_Field\_Localization\_CVPR\_2017\_paper.html**

**Key Concepts and Taxonomy:**  
This work addresses sports field localization using structured deep models that learn geometric constraints of the playing field.

**Main Contributions and Findings:**  
Combines CNNs with structured prediction to infer field boundaries and layout even in cluttered or occluded scenes.

**Limitations and Future Directions:**  
Primarily tested on soccer; requires retraining and adaptation for different sports or camera angles.

**Relevance to Primary Paper:**  
Closely tied to DeepSportradar’s camera calibration task, reinforcing the value of spatial priors and structured layout prediction in sports CV.

**[21] Peter J. Huber, 1964 – “Robust Estimation of a Location Parameter”**

**https://projecteuclid.org/journals/annals-of-mathematical-statistics/volume-35/issue-1/Robust-Estimation-of-a-Location-Parameter/10.1214/aoms/1177703732.full**

**Key Concepts and Taxonomy:**  
This foundational paper introduces robust statistical methods, particularly the Huber loss function, to reduce sensitivity to outliers in estimation problems.

**Main Contributions and Findings:**  
Defines a loss function that transitions from quadratic to linear for large residuals, balancing efficiency and robustness in optimization.

**Limitations and Future Directions:**  
Not tailored for specific modern machine learning use cases; extensions are needed for deep learning-based optimization scenarios.

**Relevance to Primary Paper:**  
DeepSportradar uses Huber loss for supervising ball diameter regression in the 3D localization task, taking advantage of its robustness to label noise and occlusions.

**[22] Paresh R. Kamble et al., 2019 – “Ball Tracking in Sports: A Survey”**

**https://ieeexplore.ieee.org/abstract/document/8718041**

**Key Concepts and Taxonomy:**  
Comprehensive review of ball tracking techniques across different sports using computer vision, including classical and deep learning methods.

**Main Contributions and Findings:**  
Highlights key challenges like motion blur, occlusion, and small object size; organizes methods by tracking approach (Kalman filters, CNNs, etc.).

**Limitations and Future Directions:**  
Limited benchmarking; lacks unified datasets or metrics for cross-comparison. Suggests the need for multi-view and physics-informed models.

**Relevance to Primary Paper:**  
Provides background for DeepSportradar’s ball 3D localization task, emphasizing the unique challenges of basketball where the ball is often partially occluded.

**[23] Lei Ke et al., 2021 – “Deep Occlusion-Aware Instance Segmentation with Overlapping Bilayers”**

**https://arxiv.org/abs/2104.08822**

**Key Concepts and Taxonomy:**  
Proposes a novel model that explicitly reasons about object occlusion using layered instance segmentation.

**Main Contributions and Findings:**  
Improves segmentation performance in crowded scenes by learning to disentangle foreground-background object layering.

**Limitations and Future Directions:**  
More computationally intensive than conventional segmentation models; performance depends on the precision of occlusion ordering.

**Relevance to Primary Paper:**  
Pertinent for DeepSportradar’s instance segmentation task where players frequently overlap, especially in key moments like rebounds or screens.

**[24] Neeraj Kumar et al., 2019 – “A Multi-Organ Nucleus Segmentation Challenge”**

[**https://ieeexplore.ieee.org/document/8918453**](https://ieeexplore.ieee.org/document/8918453)

**https://monuseg.grand-challenge.org/**

**Key Concepts and Taxonomy:**  
Medical imaging dataset and benchmark focused on accurate nucleus segmentation across different organs.

**Main Contributions and Findings:**  
Standardizes multi-domain instance segmentation in biomedical images with dense annotations and diverse modalities.

**Limitations and Future Directions:**  
Domain-specific; generalization to other contexts like sports or street scenes is limited. Variability in stain/color still poses challenges.

**Relevance to Primary Paper:**  
Provides useful precedent for designing instance segmentation benchmarks like DeepSportradar, with emphasis on annotation quality and evaluation protocols.

**[25] Yann LeCun, Yoshua Bengio, Geoffrey Hinton, 2015 – “Deep Learning”**

**https://www.nature.com/articles/nature14539**

**Key Concepts and Taxonomy:**  
A landmark review paper summarizing the principles and applications of deep learning, from supervised training to unsupervised learning and reinforcement learning.

**Main Contributions and Findings:**  
Details breakthroughs in CNNs, RNNs, and deep architectures applied across speech, vision, and NLP, sparking widespread adoption in AI.

**Limitations and Future Directions:**  
Calls for further work in generalization, interpretability, and learning from fewer labeled examples.

**Relevance to Primary Paper:**  
Serves as the theoretical backbone for all DL-based tasks in DeepSportradar, from segmentation to camera calibration to ReID.

**[26] Attila Lengyel et al., 2022 – “VIPriors 2: Visual Inductive Priors for Data-Efficient Deep Learning Challenges”**

**https://arxiv.org/abs/2208.06458**

**Key Concepts and Taxonomy:**  
Presents the VIPriors challenge series aimed at developing deep learning methods that work with limited data, leveraging visual inductive priors.

**Main Contributions and Findings:**  
Demonstrates the effectiveness of transfer learning, synthetic data, and augmentation in achieving competitive results under constrained supervision.

**Limitations and Future Directions:**  
Limited to benchmark-style evaluation without deeper exploration of domain adaptation or semi-supervised learning.

**Relevance to Primary Paper:**  
DeepSportradar's ReID dataset was introduced in VIPriors 2021, and the challenge’s emphasis on learning with limited labels is crucial for scalable deployment in sports analytics.

**[27] Wei Li et al., 2014 – “DeepReID: Deep Filter Pairing Neural Network for Person Re-Identification”**

**https://arxiv.org/abs/1409.2859**

**Key Concepts and Taxonomy:**  
Proposes one of the earliest deep learning architectures specifically tailored for person ReID using a filter pairing mechanism.

**Main Contributions and Findings:**  
Introduced a Siamese CNN that jointly learns feature representations and similarity metrics for identity matching.

**Limitations and Future Directions:**  
Struggles with background clutter and limited pose variation handling; newer part-based or attention-driven models now outperform it.

**Relevance to Primary Paper:**  
Establishes a foundation for the ReID baseline used in DeepSportradar, especially in learning discriminative embeddings for visually similar players.

**[28] Yulin Li et al., 2021 – “Diverse Part Discovery: Occluded Person Re-Identification with Part-Aware Transformer”**

**https://arxiv.org/abs/2103.13442**

**Key Concepts and Taxonomy:**  
Introduces a part-aware transformer model that improves ReID under occlusions by learning robust representations from visible body parts.

**Main Contributions and Findings:**  
Achieves state-of-the-art results on occluded ReID benchmarks by dynamically selecting informative parts for matching.

**Limitations and Future Directions:**  
Computationally expensive and reliant on part annotations or strong priors; further efficiency improvements are needed for real-time use.

**Relevance to Primary Paper:**  
DeepSportradar ReID faces similar challenges with occlusions and visual similarity among players; part-aware strategies could significantly boost performance.

**[29] Tsung-Yi Lin et al., 2014 – “Microsoft COCO: Common Objects in Context”**

[**https://arxiv.org/abs/1405.0312**](https://arxiv.org/abs/1405.0312)

**https://cocodataset.org/**

**Key Concepts and Taxonomy:**  
COCO is a large-scale dataset for object detection, segmentation, and captioning, known for its contextual richness and dense annotations.

**Main Contributions and Findings:**  
Popularized instance segmentation and set the standard for mAP-based evaluation. Encouraged algorithmic advances in detecting small and occluded objects.

**Limitations and Future Directions:**  
Mostly focused on general object categories; lacks domain-specific tasks like sports or multi-camera scenes.

**Relevance to Primary Paper:**  
DeepSportradar’s instance segmentation benchmark uses COCO’s annotation format and evaluation metrics (segm\_mAP), ensuring compatibility and standardization.

**[30] Hao Luo et al., 2019 – “Bag of Tricks and a Strong Baseline for Deep Person Re-Identification”**

[**https://arxiv.org/abs/1903.07071**](https://arxiv.org/abs/1903.07071)

**https://github.com/michuanhaohao/reid-strong-baseline**

**Key Concepts and Taxonomy:**  
Proposes a collection of practical training tricks and a strong baseline that outperforms many complex ReID models using standard architecture.

**Main Contributions and Findings:**  
Emphasizes that proper normalization, data augmentation, and loss balancing can yield high ReID accuracy without complex models.

**Limitations and Future Directions:**  
Performance gains are largely due to training improvements rather than architectural innovation; may plateau without more structural advances.

**Relevance to Primary Paper:**  
Influences the training design of DeepSportradar’s ReID baseline, which also relies on ResNet-50 and emphasizes effective training over novel architecture.

**[31] Jiaxu Miao et al., 2019 – “Pose-Guided Feature Alignment for Occluded Person Re-Identification”**

**https://arxiv.org/abs/1908.07823**

**Key Concepts and Taxonomy:**  
Introduces pose-guided alignment to improve ReID under occlusion by focusing on aligned body parts rather than global features.

**Main Contributions and Findings:**  
Improves performance on occluded ReID benchmarks by aligning features based on estimated human pose, reducing the effect of occlusion and misalignment.

**Limitations and Future Directions:**  
Requires accurate pose estimation; vulnerable to errors in crowded scenes or distorted poses. Pose estimation quality is a bottleneck.

**Relevance to Primary Paper:**  
Highly relevant for DeepSportradar’s ReID task where occlusions are common and pose-aware strategies could enhance identity discrimination.

**[32] Massimo Minervini et al., 2016 – “Finely-Grained Annotated Datasets for Image-Based Plant Phenotyping”**

**https://bmcbioinformatics.biomedcentral.com/articles/10.1186/s12859-016-1234-9**

**Key Concepts and Taxonomy:**  
Provides high-resolution datasets for instance-level plant segmentation to study fine-grained features like leaf shape and area.

**Main Contributions and Findings:**  
Demonstrates how precise, expert-annotated datasets enable detailed segmentation of biological structures for phenotypic analysis.

**Limitations and Future Directions:**  
Domain-specific; techniques don’t generalize well to dynamic or cluttered environments like sports. Primarily designed for controlled lab conditions.

**Relevance to Primary Paper:**  
Inspires the meticulous annotation process in DeepSportradar’s segmentation task, especially where object boundaries and fine detail are crucial.

**[33] Davy Neven et al., 2019 – “Instance Segmentation by Jointly Optimizing Spatial Embeddings and Clustering Bandwidth”**

**https://arxiv.org/abs/1805.04554**

**Key Concepts and Taxonomy:**  
Proposes a bottom-up instance segmentation method using pixel embeddings and dynamic clustering to form object instances.

**Main Contributions and Findings:**  
Achieves competitive results with lightweight models; avoids reliance on bounding box proposals by learning pixel affinities directly.

**Limitations and Future Directions:**  
Clustering step can be sensitive to hyperparameters; challenging to tune for real-time inference or high-resolution images.

**Relevance to Primary Paper:**  
Provides a viable bottom-up alternative for DeepSportradar’s instance segmentation task, particularly in dense or occluded scenes.

**[34] Niels Sayez & Christophe De Vleeschouwer, 2022 – “Accelerating the Creation of Instance Segmentation Training Sets Through Bounding Box Annotation”**

**https://arxiv.org/abs/1805.04554**

**Key Concepts and Taxonomy:**  
Describes a semi-automated pipeline for generating instance masks from simpler bounding box annotations using interactive tools and models.

**Main Contributions and Findings:**  
Enables rapid dataset creation with high mask accuracy, significantly reducing annotation time while maintaining quality.

**Limitations and Future Directions:**  
Requires careful user supervision and post-processing to avoid bias or systematic errors. May not generalize to complex scenes without retraining.

**Relevance to Primary Paper:**  
Used directly in DeepSportradar to generate high-quality human instance masks for player segmentation, accelerating the annotation pipeline.

**[35] Long Sha et al., 2020 – “End-to-End Camera Calibration for Broadcast Videos”**

**https://arxiv.org/abs/2007.03367**

**Key Concepts and Taxonomy:**  
Presents a deep learning pipeline for predicting full camera calibration parameters from broadcast sports footage using court markings.

**Main Contributions and Findings:**  
Achieves accurate calibration by predicting court line segmentation maps and estimating the camera matrix via geometric fitting.

**Limitations and Future Directions:**  
Accuracy can degrade with partial occlusions or non-standard camera angles; assumes known field geometry and consistent lighting.

**Relevance to Primary Paper:**  
Directly informs DeepSportradar’s camera calibration task; similar approach using line segmentation and projection matrix estimation with a CNN-based backbone.

**[36] Karen Simonyan & Andrew Zisserman, 2015 – “Very Deep Convolutional Networks for Large-Scale Image Recognition” (VGGNet)**

**https://arxiv.org/abs/1409.1556**

**Key Concepts and Taxonomy:**  
Introduces the VGG family of CNNs with deep but uniform architecture using 3×3 convolutional layers, emphasizing depth over complexity.

**Main Contributions and Findings:**  
VGG models achieved top results on ImageNet and served as a foundation for many subsequent CV architectures due to their simplicity and transferability.

**Limitations and Future Directions:**  
Large number of parameters leads to high memory and computation cost; later models (ResNet, MobileNet) offer similar or better performance with improved efficiency.

**Relevance to Primary Paper:**  
VGG16 is used as the baseline architecture in DeepSportradar for the 3D ball localization task, chosen for its simplicity and effectiveness in regression problems.

**[37] Yifan Sun et al., 2018 – “Beyond Part Models: Person Retrieval with Refined Part Pooling”**

**https://arxiv.org/abs/1711.09349**

**Key Concepts and Taxonomy:**  
Proposes a part-based model for person ReID that learns features from spatially consistent body parts without requiring pose estimation.

**Main Contributions and Findings:**  
Demonstrates that partitioning the person image into parts and learning refined features for each improves ReID accuracy over global-only embeddings.

**Limitations and Future Directions:**  
Performance may still degrade in extreme occlusion or pose variation cases; combining with attention or pose-guided alignment could further improve results.

**Relevance to Primary Paper:**  
Informative for improving DeepSportradar’s ReID task, especially in scenarios where uniform player clothing requires part-based disambiguation.

**[38] Gabriel Van Zandycke, 2021 – “DeepSport Dataset”**

**https://deepsport.dvl.in.tum.de/**

**Key Concepts and Taxonomy:**  
Initial version of the DeepSport dataset focused on ball 3D localization in basketball games, offering synchronized multi-view imagery and calibration data.

**Main Contributions and Findings:**  
Provided raw high-resolution frames and precise annotations, forming the basis for real-world CV tasks in sports analytics.

**Limitations and Future Directions:**  
Early release only included ball annotations; lacked segmentation and ReID labels, which were added in later extensions.

**Relevance to Primary Paper:**  
Forms the core dataset behind DeepSportradar, later expanded to support four major CV tasks, improving realism and task diversity.

**[39] Gabriel Van Zandycke & Christophe De Vleeschouwer, 2022 – “3D Ball Localization from a Single Calibrated Image”**

**https://deepsport.dvl.in.tum.de/**

**Key Concepts and Taxonomy:**  
Presents a method to estimate ball 3D location from a single image using its apparent diameter and known camera calibration.

**Main Contributions and Findings:**  
Combines image-based regression with geometric constraints to achieve accurate ball localization in the 3D space using monocular imagery.

**Limitations and Future Directions:**  
Accuracy can be affected by occlusions, ball motion blur, or annotation noise. Future work may include temporal filtering or multi-view integration.

**Relevance to Primary Paper:**  
Provides the baseline for the 3D ball localization task in DeepSportradar, including dataset setup, model architecture, and performance metrics.