# State of the art on real-time pose estimation - Application to the case of surgery for the prevention of nosocomial diseases

Hamza Dribine[1], Abdellatif Belmady[1], Fatine Boussatine[1], Salma Khmassi[1], Mohammed El Rhabi[1], Phillipe Lutman[2]

[1]Ecole Centrale Casablanca
[2]REMED-IA

# Context and Motivation

Out of every 100 patients in acute-care hospitals, seven patients in high-income countries and 15 patients in low- and middle-income countries will acquire at least one health care-associated infection (HAI) during their hospital stay.

On average, **1 in every 10 affected patients will die from their HAI** (1).

One of the highest prevalence of HAIs in hospital wards was related to the surgery wards with the prevalence rate of **0.43** (2).

It was estimated that 3% of surgical procedures performed in 2010 in France resulted in infection, resulting in an annual cost of **€57 892 715** (3).

| | | |
|---|---|---|
| AFRO | 94 | 0.270 |
| AMRO | 18 | 0.096 |
| EMRO | 103 | 0.125 |
| EURO | 114 | 0.114 |
| SEARO | 24 | 0.129 |
| WPRO | 47 | 0.097 |

**Meta-analysis based on WHO regions.**

(1) WHO 2022 Global report
(2) Global prevalence of nosocomial infection: A systematic review and meta-analysis, Samira Raoofi (2023)
(3) Evaluating the clinical and economic burden of healthcare-associated infections during hospitalizations for surgery in France, L.LAMARSALLE (2013)

# Complexity of the operating environment

The application of motion detection models in the modern operating room faces a number of challenges :

- High visual complexity
- Occlusion and clutter from various equipment
- Loose clothes worn by clinicians
- Proximity of the people present in the scene

# Project Framework

**Remed-IA** aims to develop an automated system that can aid circulating nurses in ensuring the proper adherence to operating room protocols, thereby reducing the likelihood of aseptic errors.

The depth information provided by RGB-D cameras is highly suitable for dealing with occlusion and clustering complexities.

**RGB-D Cameras**

**Input**

All Encompassing Model

**Output**


**Protocol Breach Warning**

Our project aims to identify the most effective pose estimation model in the state-of-the-art that is suitable for use in an operating room setting.

# Pose Estimation

Human pose estimation aims to predict the poses of human body parts and joints in images or videos. Since pose motions are often driven by some specific human actions, knowing the body pose of a human is critical for action recognition and video understanding.


**3D vs 2D**

**Multi Person**

**Open Source**

**Bottom-up vs Top-down**

# Pose estimation state of the art

Through a review of the state of the art, we have identified 6 open-source multi-person pose estimation models that excel in handling occlusion and cluster complexities :

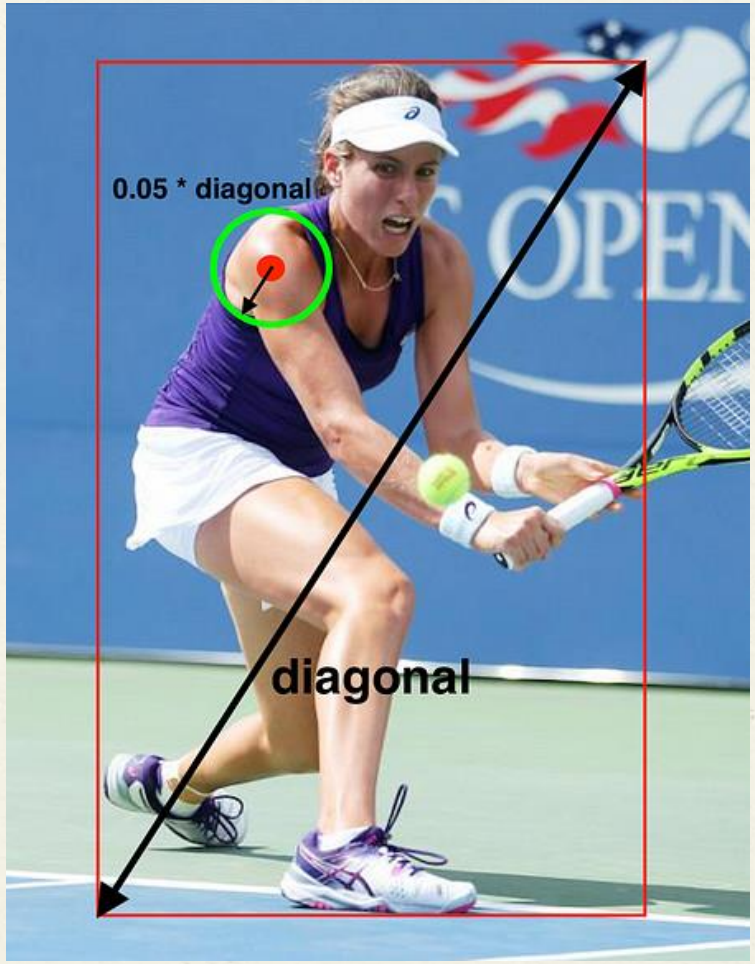

- YoloV7 (Chien-Yao Wang , Alexey Bochkovskiy, and Hong-Yuan Mark Liao. (2022). YOLOv7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors)
- OpenPose (Cao, Z., Simon, T., Wei, S. E., & Sheikh, Y. (2017). Realtime multi-person 2D pose estimation using part affinity fields. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 7291-7299).)
- MoveNet (Poon, A., Zhu, T., Sudarshan, M., & Chen, Z. (2021). MoveNet: Efficient Convolutional Neural Networks for Mobile Vision Applications. arXiv preprint arXiv:2102.04664.)
- PoseNet (Kendall, A., Grimes, M., & Cipolla, R. (2015). PoseNet: A convolutional network for real-time 6-DOF camera relocalization. In Proceedings of the IEEE International Conference on Computer Vision (pp. 2938-2946).)
- DeeperCut (Newell, A., Yang, K., & Deng, J. (2016). Stacked hourglass networks for human pose estimation. In European Conference on Computer Vision (pp. 483-499).)
- DCPose (Tang, W., Xiao, B., Yan, S., & Wang, X. (2018). DCPose: Dynamic Curriculum Learning for 3D

# Evaluation Method

We evaluated our models by using the Percentage of Detected Joints (PDJ) metric, in which the Detected Joint is considered correct if the distance between the predicted and the true joint is within a certain fraction of the bounding box diagonal.

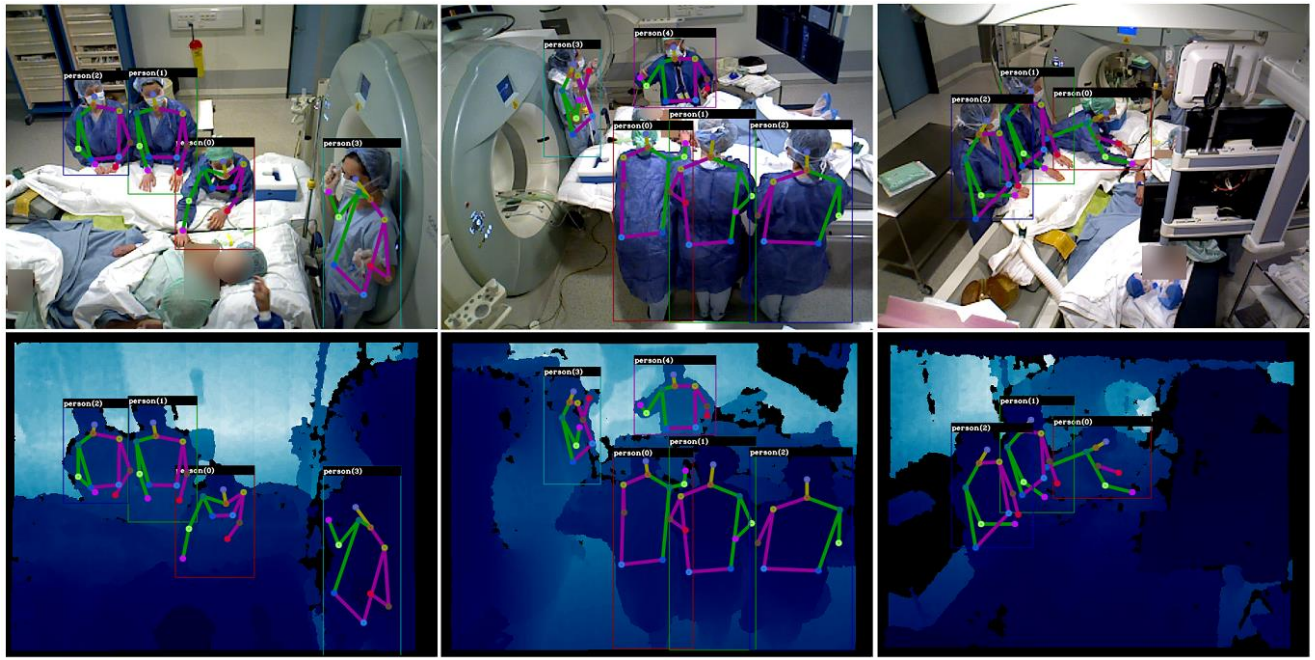
$$PDJ = \frac{\sum_{i=1}^{n} bool(d_i < 0.05 * diagonal)}{n}$$

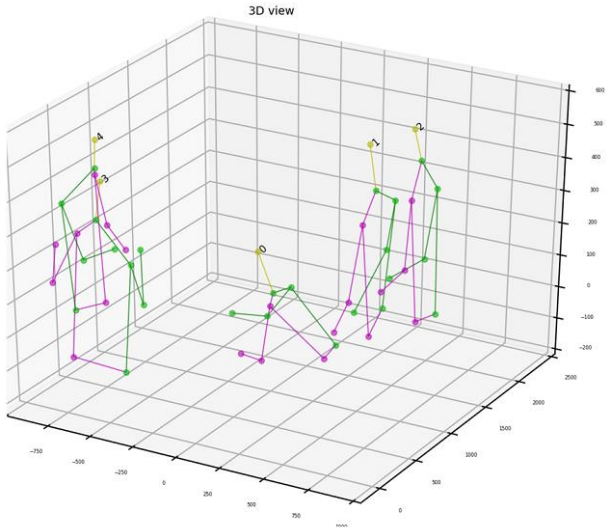
- $di$ — the euclidian distance between ground truth keypoint and predicted keypoint;
- $bool(condition)$ — a function that returns 1 if the condition is true, 0 if it is false;
- $n$ —the number of key points on the image.



# Operating Room Database

We decided to utilize the Multi-view RGB-D Operating Room Dataset1 (MVOR) for our comparative assessment, as it is the first publicly available dataset captured during genuine medical procedures.
The dataset encompasses 732 multi-view frames that are synchronized and captured by three RGB-D cameras situated in a hybrid OR. The ground truth annotations consist of 4699 human bounding boxes, 2926 2D upper-body poses and 1061 3D upper-body poses.



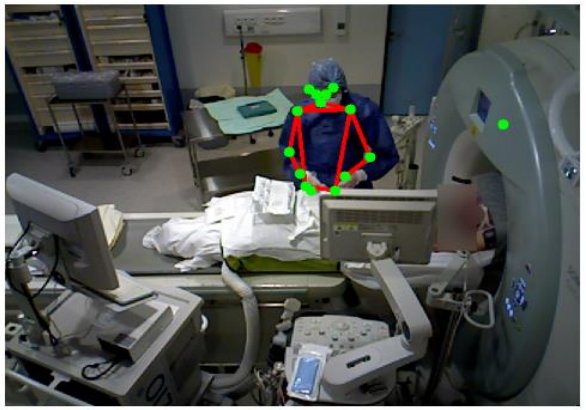



[1] Vinkle Srivastav, Thibaut Issenhuth, Abdolrahim Kadkhodamohammadi, Michel de Mathelin, Afshin Gangi, and Nicolas Padoy. (2021). MVOR: A Multi-view RGB-D Operating Room Dataset for 2D and 3D Human Pose Estimation

# Results

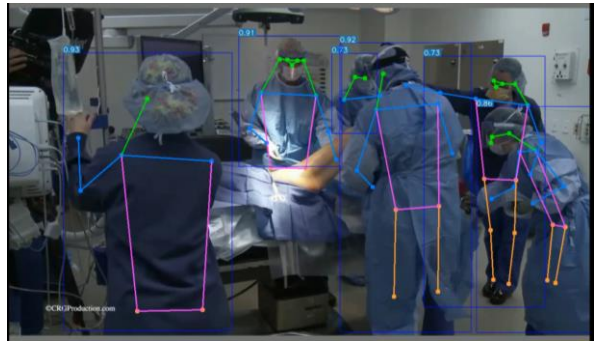
*Fraction = 0,05*

| PDJ by body-part | MoveNet | PoseNet | OpenPose | DCPose | DeepCut | YoloV7 |
|---|---|---|---|---|---|---|
| **Head** | 64,3 | 42,1 | 71,2 | 70,1 | 67,8 | 86,5 |
| **Shoulders** | 61,2 | 43,5 | 72,3 | 68,2 | 63,1 | 89,1 |
| **Elbows** | 68,1 | 47,1 | 78,1 | 69,11 | 68,1 | 90,2 |
| **Wrists** | 70,2 | 49,2 | 80,2 | 67,1 | 69,2 | 86,2 |
| **Hips** | 54,1 | 57,2 | 85,1 | 72,1 | 67,2 | 84,1 |
| **Knees** | 68,4 | 60,1 | 78,1 | 54,1 | 72,1 | 92,1 |
| **Ankles** | 54,1 | 47,5 | 81,2 | 70,4 | 73,5 | 84,2 |
| **Mean** | 62,9 | 49,5 | 78,0 | 67,3 | 68,7 | 87,4 |



**MoveNet**

**OpenPose**



**YoloV7**

The **YOLOV7** model exhibited the highest level of performance, surpassing all other models, with an average accuracy of correctly predicted joints at **87.4%**.