visuAAL

Privacy-Aware and Acceptable Video-Based Technologies and Services for Active and Assisted Living

Context Recognition for the Application of Visual Privacy

ESR 14

University of Alicante
20/4/2023

Kooshan Hashemifard

This project has received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No 861091".

Universitat d'Alacant
Universidad de Alicante
Project Coordinator

RWTH AACHEN UNIVERSITY

Stockholm University

Trinity College Dublin
Coláiste na Tríonóide, Baile Átha Cliath
The University of Dublin

TU WIEN

Introduction

- Demographic changes
- Burden to care personnel and facilities
- Damage to autonomy, self-esteem and spirit
- Ambient-assisted living (AAL) and sensors
- Video-based technology
- The most directed and natural way to record events
- Provide richer information
- Easy to interpret by unauthorized viewers and privacy issues

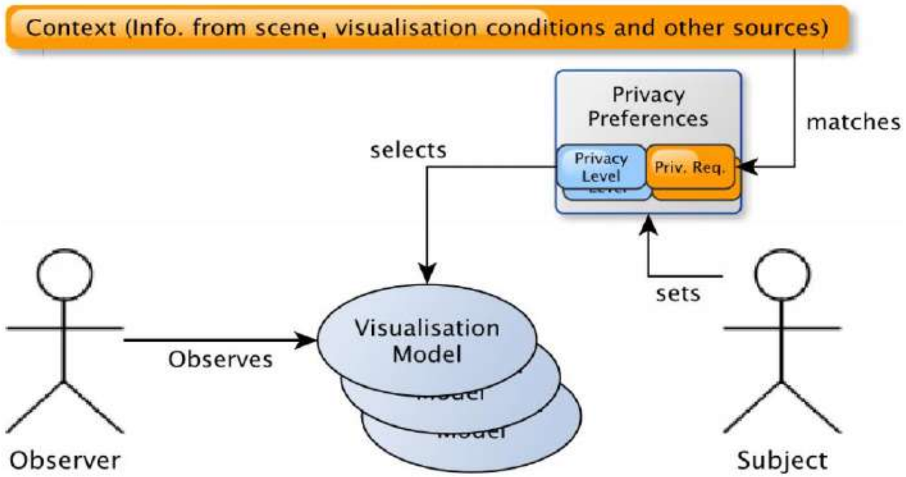
Research project: Context recognition for the application of visual privacy

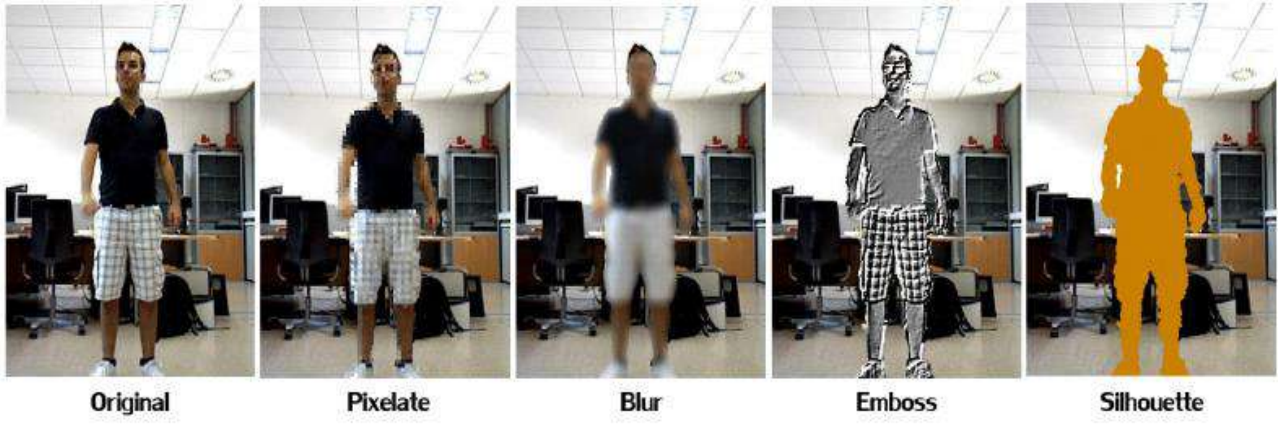
- Balance between privacy-intelligibility in video
- Can we find a method to make the balance? Understand what is happening and preserve privacy
- Previous work and privacy-by-context
- Level-based visualisation, selected according to the context
- Privacy is subjective

Introduction

Research project: Context recognition for the application of visual privacy

- Estimate the variables automatically to be used along with preferences for different visualisation
- Extract rich information from the context to empower people to adapt privacy
- Computer vision algorithms for continuous estimation of the context
- context can be defined by variables: Appearance, Activity, Event, Place, ...

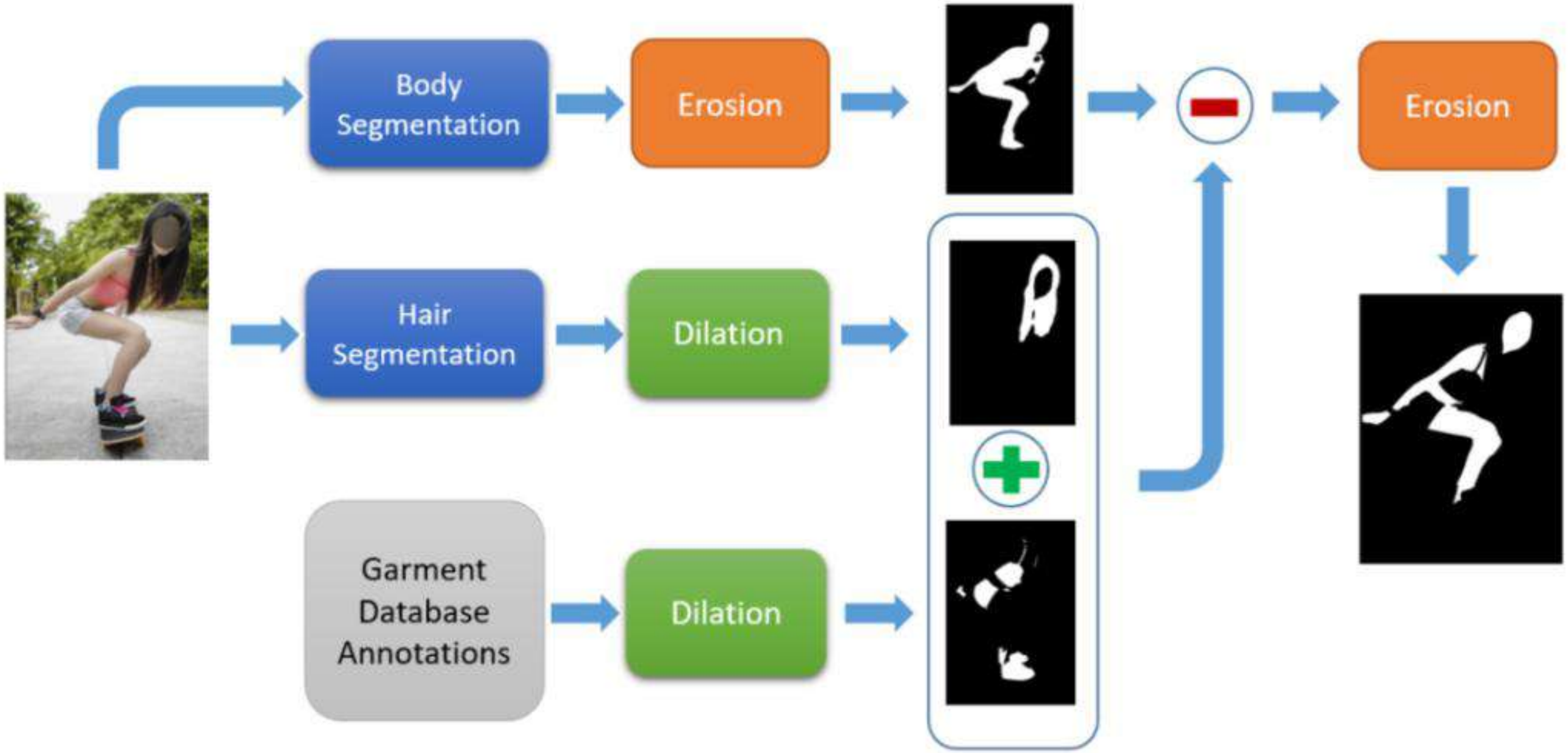




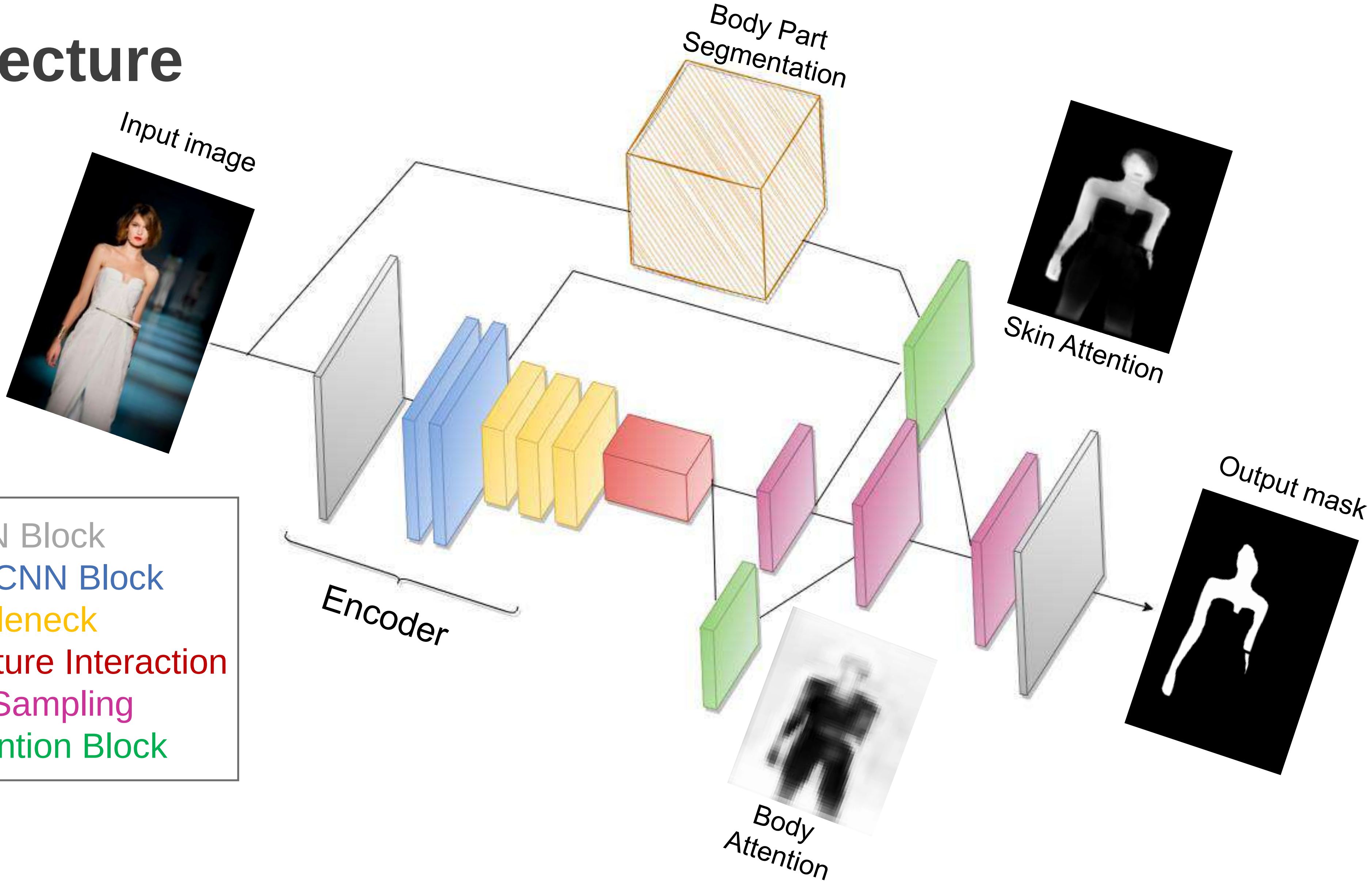
Appearance Detection

- Serious concerns about person appearance recorded in a video
- Appearance detection can be interpreted as nudity detection in the context of AAL
- Skin Detection plays a crucial role in nudity detection
- Estimate the degree of nudity

1- Dataset


Body Segmentation
Erosion
Hair Segmentation
Dilation
Garment Database Annotations
Dilation
Erosion

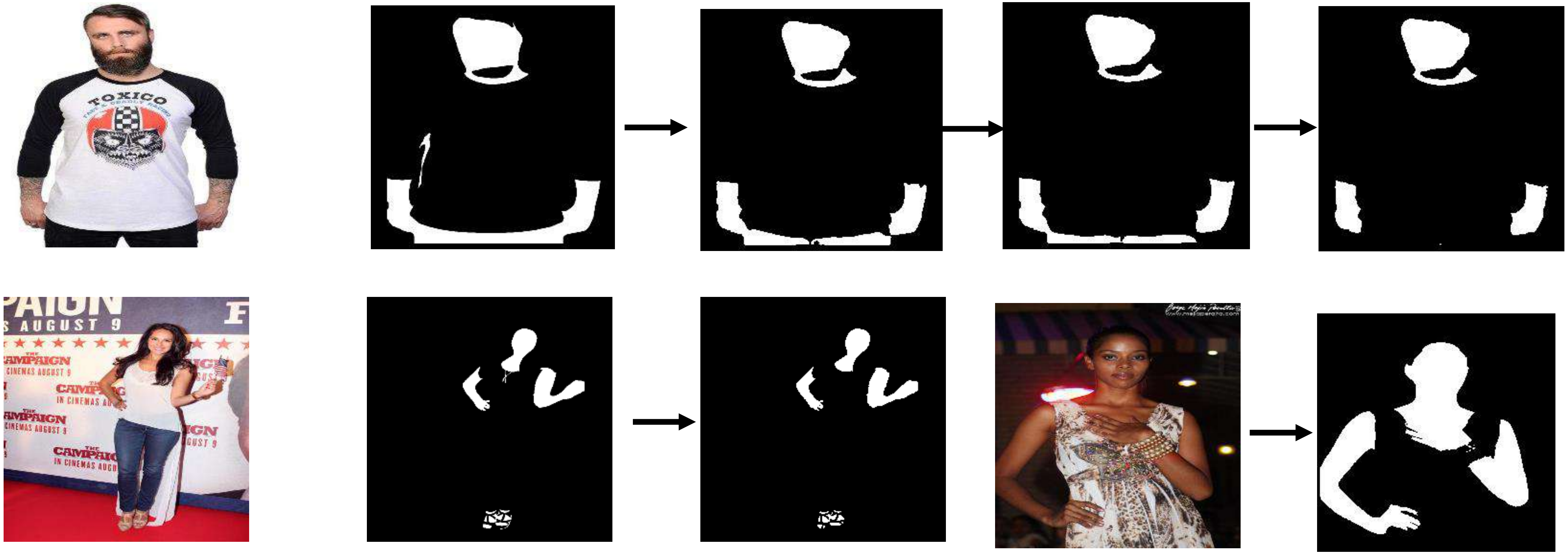
2- Architecture



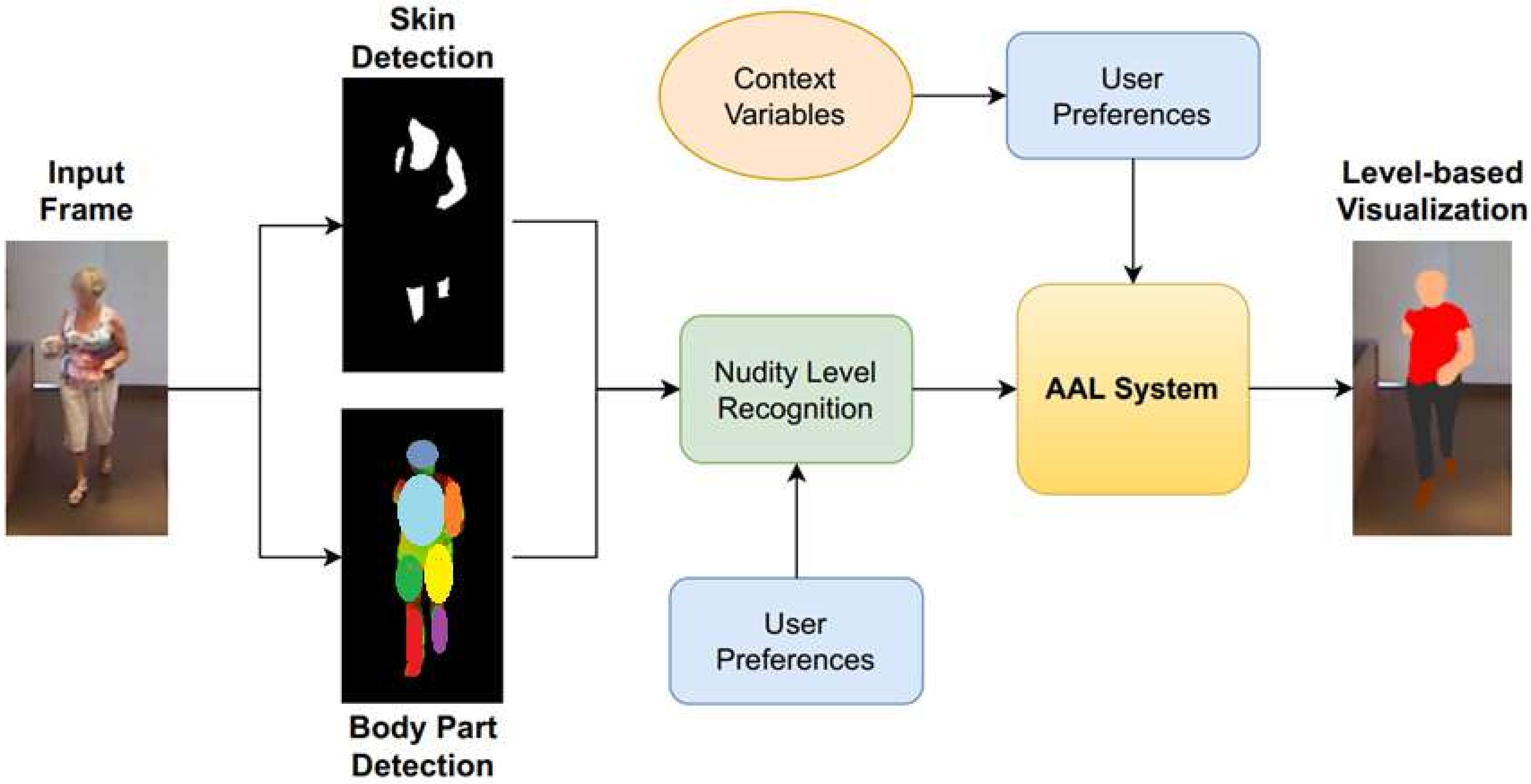
CNN Block
DWCNN Block
Bottleneck
Feature Interaction
Up Sampling
Attention Block

Recap

3- Training

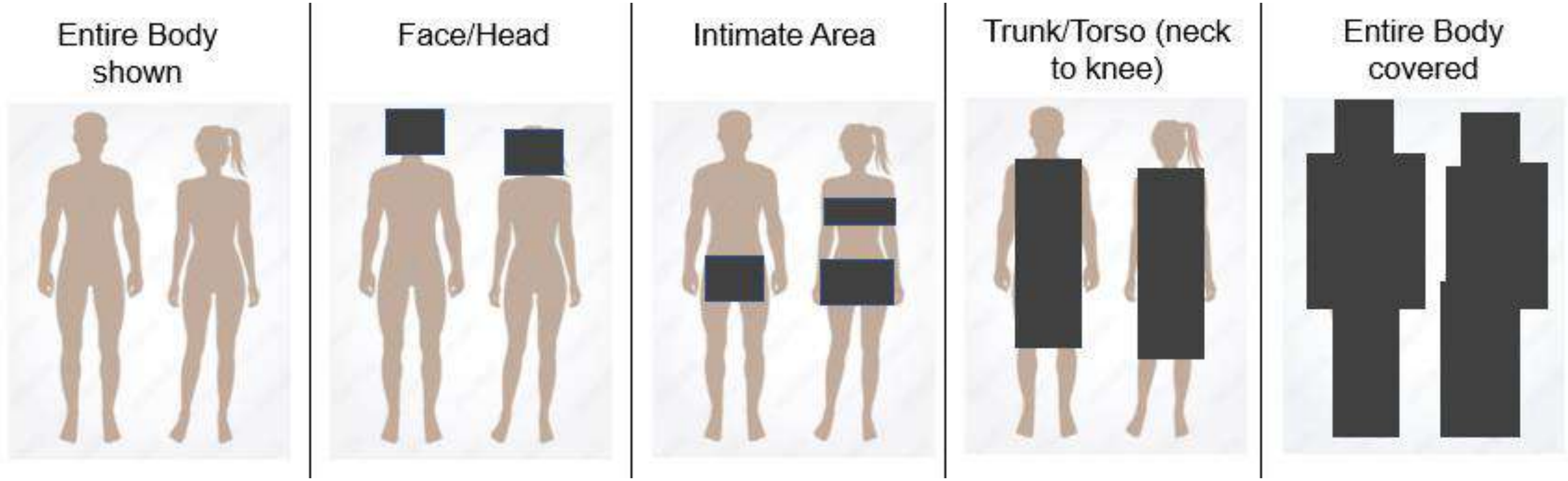


4- Nudity Recognition



Nudity Level

Nudity Level	Description
1	*Completely covered*
2	*Covered torso (neck to knee)*
3	*Covered intimate areas*
4	*Covered faces*
5	*Full body or exposed intimate areas*

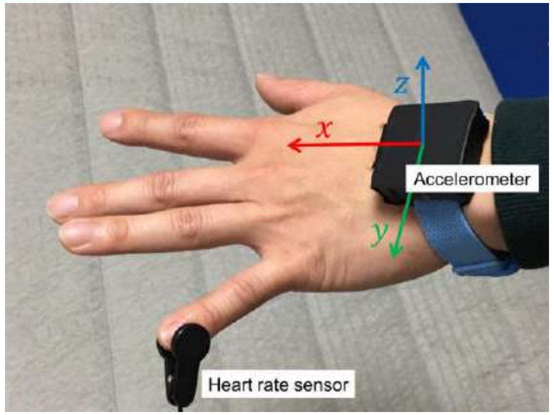


Fall Detection

- Millions of 65 and older people, fall each year at least once
- Smart devices and sensors are used to assist independent living
- In AAL environments, mobile robots can play a crucial role

Fall Detection

Fall Detection Approaches



- Wearables:
  - Pros: generally accurate and reliable, privacy preserving
  - Cons: should be worn all the time
- Video-based:
  - Pros: non invasive, passive monitoring
  - Cons: privacy concerns, exposure to cameras' sight

Assistive Mobile Robots

One Potential Solution: Assistive Mobile Robots

Growing interest in assistive robots for monitoring and aiding older people in AAL environments

Advantages:

- Covers all areas
- Possible processing on the edge for privacy

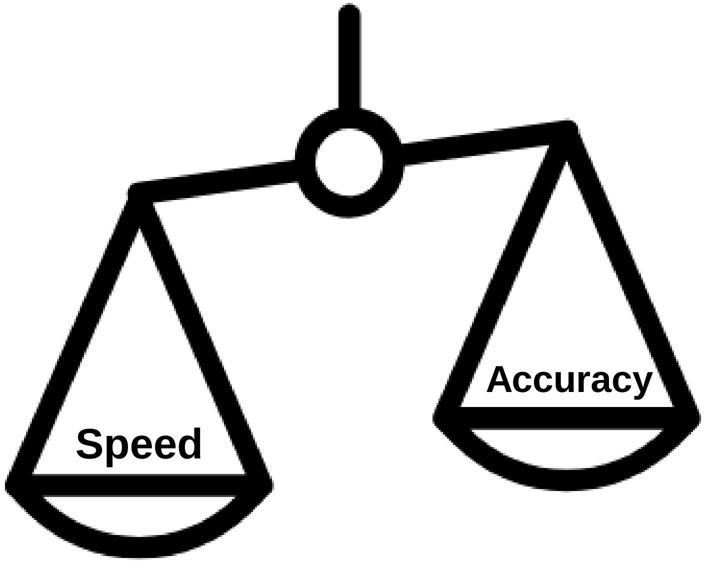
Edge-AI Device

1st Secondment at Maynooth University

- Fall Detection using Privacy-preserving Edge-AI Luxonis OAK-D Camera
- A promising approach to privacy-preserving image processing since it allows for local processing of data
- Only high-level information is transmitted across the network
- Relatively powerful considering size and price

Challenges

- Not only accuracy is highly critical, but also needs to quickly detect a fallen person
- Although more affordable, limitations in terms of computational resource
- A single device may be responsible for managing multiple concurrent tasks



visuAAL

Fall Detection on Edge-AI

Steps for the reliable method:

1. Selection of the Backbone Network and Pre-Trained Model
2. Transfer Learning for Fallen Person Detection
3. Knowledge Distillation for Efficiency Improvement
4. Deploying on Luxonis Camera
5. Evaluation In-the-wild with Luxonis

Backbone Network

- Various models and architectures for general object detection, trained on large datasets such as ImageNet and COCO
- Evaluated and compared based on metrics such as speed, mean average precision, and output type
- Although can recognize the "person" class, their accuracy evaluations may not be entirely reliable because detecting fallen individuals is our primary focus
- fallen individuals can assume various positions, such as twists, fetal positions, or abnormal poses, which may not be commonly represented in normal human detection datasets, leading to lower detection rates for these poses

Backbone Network

Table 1: Initial performance evaluation of pre-trained models over detecting the "person" class in Elderly set.

Model	Detection Precision	Detection Recall	F1-score	Average IoU
MobileNet SSD	84.02%	71.04%	76.98%	72.92%
EfficientDet	88.64%	76.32%	82.02%	73.41%
CenterNet	83.33%	51.53%	63.68%	**73.57%**
YOLOv6 S	90.61%	74.22%	81.60%	72.72%
YOLOv6 M	**95.50%**	72.35%	82.32%	72.43%
YOLOv6 L	92.50%	**85.71%**	**88.97%**	72.82%

Transfer Learning

- End-to-end approach with 2 classes: fallen and upright individuals
- Top layer of the object detector network was replaced with two output neurons for each status
- Pre-trained parameters retained, except for the last layer
- Fine-tune on E-FPDS dataset, contains 6982 images captured in indoor environments, with 5023 instances of falls and 2275 instances of non-falls in various scenarios, including variations in pose, size, occlusions, and lighting

Transfer Learning

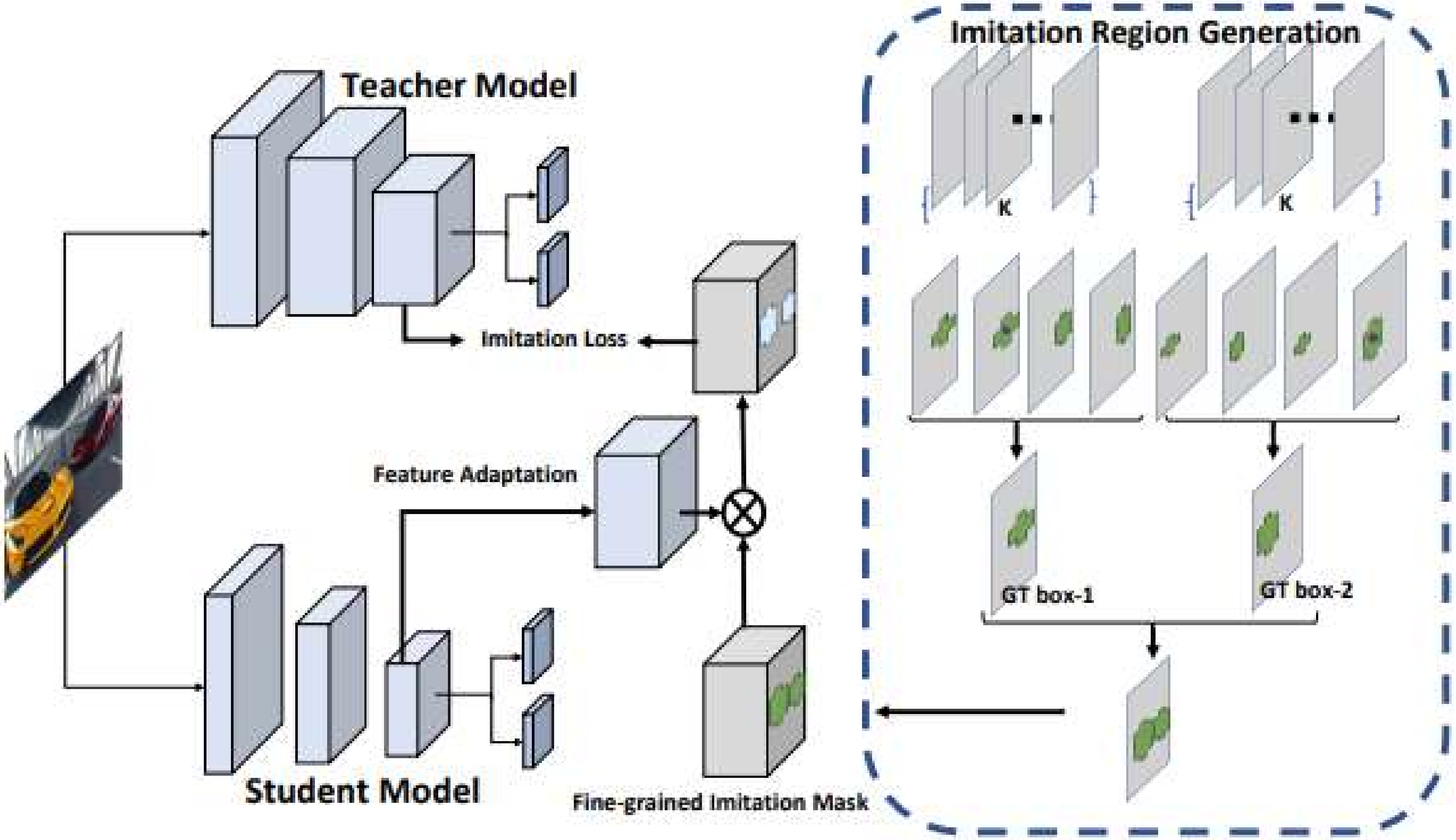
Table 2: Fallen person detection performance on Elderly set by fine-tuned models.

Model	Precision	Recall	F1-score	Average IoU
MobileNet SSD	85.41%	86.52%	85.96%	49.25%
EfficientDet	75.28%	**94.88%**	83.95%	53.28%
CenterNet	92.60%	92.50%	92.54%	67.75%
YOLOv6 S	85.30%	88.5%	86.87%	66.49%
YOLOv6 M	88.10%	86.50%	87.29%	70.41%
YOLOv6 L	**98.42%**	92.25%	**95.23%**	**71.44%**

Knowledge Distillation

- state-of-the-art CNN-based networks can be computationally expensive and difficult to deploy on smaller devices, especially in real-time and multi-tasking scenarios
- knowledge distillation has emerged to directly learn compact models by transferring knowledge from a large model (teacher network) to a smaller one (student network)
- Reducing computational costs without sacrificing validity
- Fine-grained Feature Imitation method
- local features in the object region and near its anchor location contain important information and are more crucial for the detector and how the teacher model tends to generalize
- The smaller student detector is trained by using both ground truth supervision and feature response imitation on object anchor locations from the teacher networks

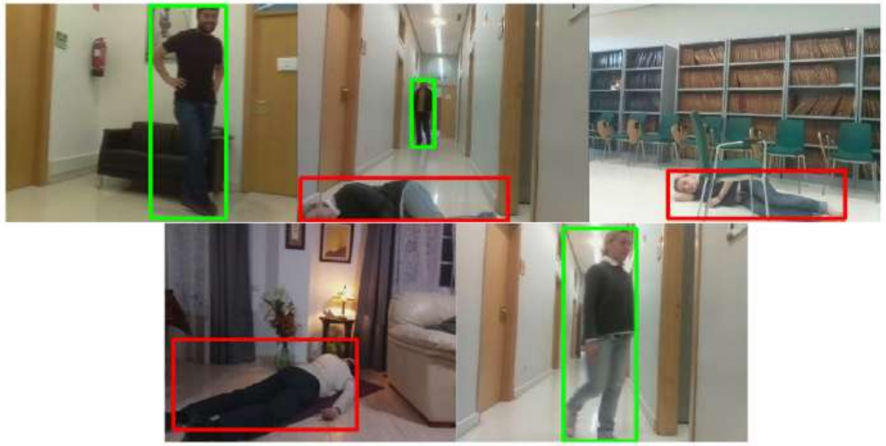
Knowledge Distillation



Knowledge Distillation

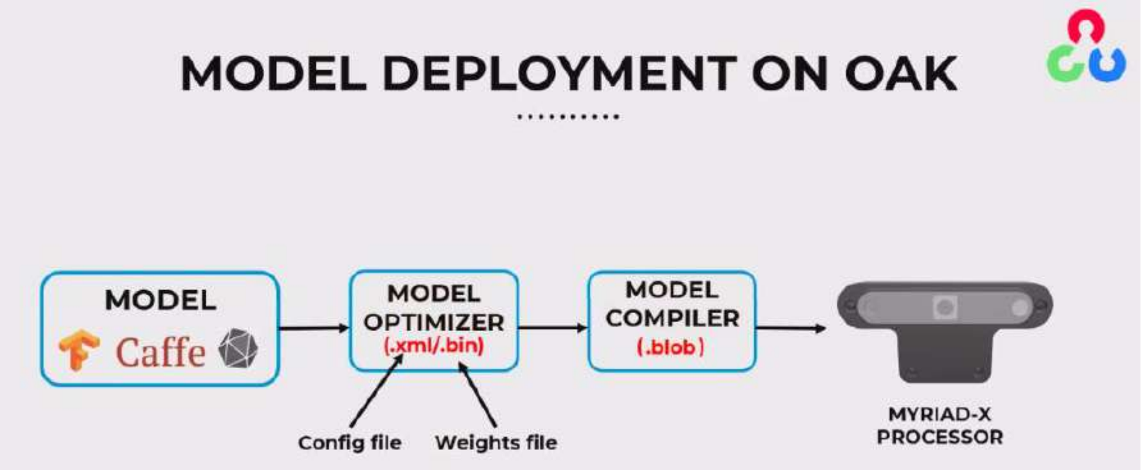
Table 3: Fallen person detection performance of YOLO versions on E-FDPS test set and paramter numbers.

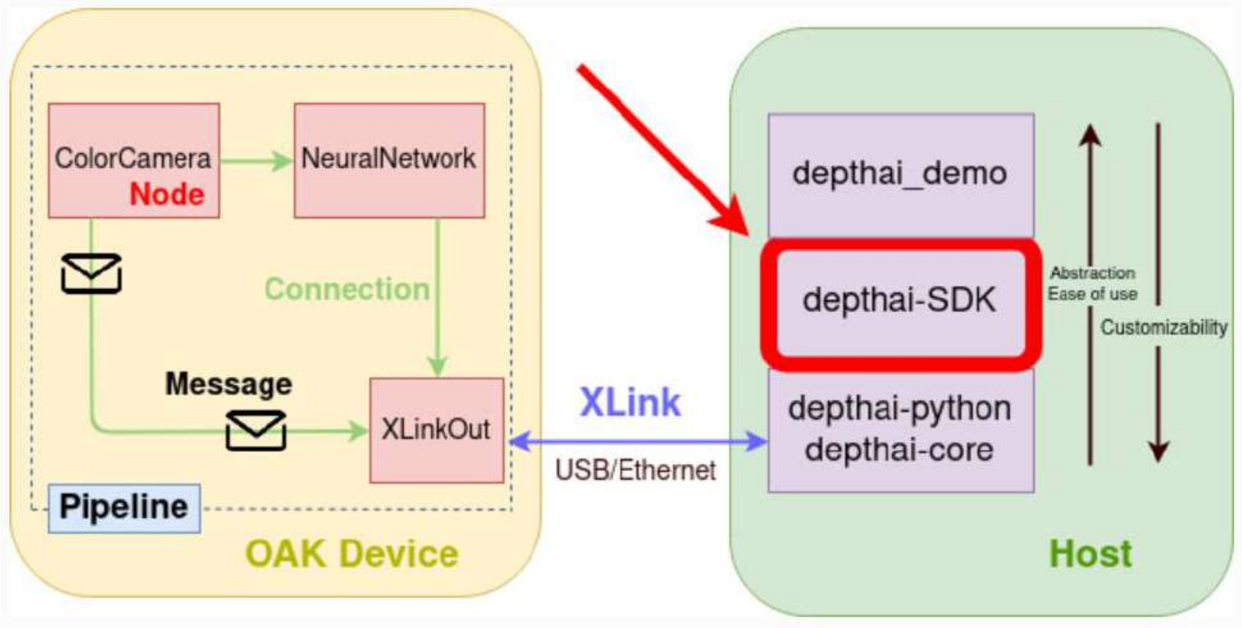
Model	Precision	Recall	mAP50	mAP50-95	Parameter Number
YOLOv6 L	98.02%	98.25%	98.31%	61.84%	59.6M
YOLOv6 M	96.55%	96.63%	97.41%	62.86%	34.9M
YOLOv6 S	95.71%	96.78%	97.32%	60.04%	18.5M
YOLOv5 L	97.67%	96.29%	98.65%	62.91%	46.5M
YOLOv5 S	93.91%	91.17%	95.11%	56.80%	7.2M
YOLOv5 S-v6 L	96.52%	95.10%	97.22%	61.52%	7.2M



Deploying on Luxonis Camera

- Models cannot be used directly by DepthAI
- First, converted to the OpenVINO IR
- Then, the IR model is compiled into a MyriadX blob
- layers and loss functions should be compatible and supported by OpenVINO





visuAAL

Evaluation In-the-wild with Luxonis

- model's performance in real-world settings after conversions
- videos, captured using the Luxonis OAK-D camera
- 38 videos recorded from various angles and labeled

Evaluation In-the-wild with Luxonis

Table 4: Fallen person detection performance of YOLO versions on in-the-wild set and the FPS on camera.

Model	Precision	Recall	mAP50	mAP50-95	FPS
YOLOv6 L	99.12%	98.64%	99.45%	65.58%	1
YOLOv6 M	94.95%	98.21%	98.64%	65.33%	5
YOLOv6 S	92.71%	95.22%	98.10%	62.67%	10
YOLOv5 L	93.48%	88.36%	94.89%	58.12%	1
YOLOv5 S	83.67%	86.44%	93.37%	46.77%	15
YOLOv5 S-v6 L	94.46%	91.83%	95.04%	57.29%	15

Other Activities

- Other Activities that have been carried out:
- Visualization demo real-time
- Changing visualization according to fall detection
- Optimizing and porting to another edge-AI device, Jetson Xavier NX

Publications

- Underneath Your Clothes: A Social and Technological Perspective on Nudity in The Context of AAL Technology (Petra 2022)
- From Garment to Skin: The visuAAL Skin Segmentation Dataset (ICIAP 2022)
- Weakly Supervised Human Skin Segmentation using Guidance Attention Mechanisms (Submitted, Image and Vision Computing)
- A Fallen Person Detector with a Privacy-preserving Edge-AI Camera (ICT4AWE)
- Appearance Detection on visual skin dataset
- ODIN: An OmniDirectional INdoor dataset

Publications

From Garment to Skin: The visuAAL Skin Segmentation Dataset *

Kooshan Hashemifard[0000−0001−5086−3064] and Francisco Florez-Revuelta[0000−0002−3391−711X]

Department of Computing Technology, University of Alicante, San Vicente del Raspeig, 03690, Spain
{k.hashemifard,francisco.florez}@ua.es

Abstract. Human skin detection has been remarkably incorporated in different computer vision and biometric systems. It has been receiving increasing attention in face analysis, human tracking and recognition, and medical image analysis. For many human-related recognition tasks, using skin detection cue could be a proper choice. Despite the vast area of usage and applications for skin detection, not many large or reliable skin detection datasets are available, and many of the existing ones, are originally created for other tasks such as hand tracking or face analysis. In this paper, we propose a methodology for extracting skin pixels from garment segmentation and recognition datasets. This is achieved by using deep learning methods to generate automatic skin label masks from them by exploiting human body and hair segmentation and provided garment masks. Following this approach, a large human skin segmentation dataset is introduced. A validation set is also manually segmented in order to evaluate the accuracy of the output skin masks. Finally, usual methods for skin detection and segmentation are evaluated on this new dataset.

Keywords: Skin segmentation · Dataset.

Underneath Your Clothes: A Social and Technological Perspective on Nudity in The Context of AAL Technology

Caterina Maidhof*
Kooshan Hashemifard*
maidhof@comm.rwth-aachen.de
k.hashemifard@ua.es
RWTH Aachen University
Aachen, Aachen, Germany
University of Alicante
San Vicente Del Raspeig, Alicante, Spain

Julia Offermann
RWTH Aachen University
Aachen, Aachen, Germany
offermann@comm.rwth-aachen.de

Martina Ziefle
RWTH Aachen University
Aachen, Aachen, Germany
ziefle@comm.rwth-aachen.de

Francisco Florez-Revuelta
University of Alicante
San Vicente Del Raspeig, Alicante, Spain
francisco.florez@ua.es

ABSTRACT

One promising way to tackle healthcare challenges due to demographic change lies in the development of user-tailored AAL technologies. Video-based AAL technologies have the potential to provide rich information - in particular about accidents such as falls. However, as visual AAL is designed to record some parts of daily life at home, privacy concerns may comprise recordings in unwanted appearances and especially while being nude. Here, collaborative research is necessary to enable the development of user-tailored (visual) AAL technologies taking into account future users' needs and concerns. This article presents an interdisciplinary collaboration investigating perceptions of nudity from a social perspective, and developing solutions on nudity detection from a technical perspective. Focusing on first empirical insights and a proposed methodology for level-based nudity detection, this article concludes with interdisciplinary learnings, derived guidelines, and implications for future collaborative research.

ACM Reference Format:
Caterina Maidhof, Kooshan Hashemifard, Julia Offermann, Martina Ziefle, and Francisco Florez-Revuelta. 2022. Underneath Your Clothes: A Social and Technological Perspective on Nudity in The Context of AAL Technology. In *The15th International Conference on PErvasive Technologies Related to Assistive Environments (PETRA '22), June 29-July 1, 2022, Corfu, Greece.* ACM, New York, NY, USA, 7 pages. https://doi.org/10.1145/3529190.3534733

1 INTRODUCTION

One of the promising ways to tackle healthcare challenges due to ageing societies and lacking care personnel lies in the development of user-tailored Ambient Assisted Living (AAL) technologies. AAL technologies and systems are intended to be a constant part of the everyday life of older people (in need of care) and aim to improve people's life quality, well-being, autonomy, and safety (e.g., [3, 18]) which also intends to relieve professional and familial caregivers in their care burdens [29]. Taking advantage of Information and Communication Technologies (ICT), different sensors and artificial

A Fallen Person Detector with a Privacy-preserving Edge-AI Camera

Kooshan Hashemifard[1][a], Francisco Florez-Revuelta[1][b] and Gerard Lacey[2][c]

[1] *Department of Computing Technology, University of Alicante, San Vicente del Raspeig, Spain*

[2] *Department of Electronic Engineering, Maynooth University, Maynooth, Ireland*

{k.hashemifard, francisco.florez}@ua.es, gerry.lacey@mu.ie

Keywords: Ambient-Assisted Living (AAL), Privacy-preserving Camera, Fallen Person Detection, Edge-AI

Abstract: As the population ages, Ambient-Assisted Living (AAL) environments are increasingly used to support older individuals' safety and autonomy. In this study, we propose a low-cost, privacy-preserving sensor system integrated with mobile robots to enhance fall detection in AAL environments. We utilized the Luxonis OAK-D Edge-AI camera mounted on a mobile robot to detect fallen individuals. The system was trained using YOLOv6 network on the E-FPDS dataset and optimized with a knowledge distillation approach onto the more compact YOLOv5 network, which was deployed on the camera. We evaluated the system's performance using a custom dataset captured with a robot-mounted camera. We achieved a precision of 96.52%, a recall of 95.10%, and a recognition rate of 15 frames per second. The proposed system enhances the safety and autonomy of older individuals by enabling the rapid detection and response to falls.

ODIN: An OmniDirectional INdoor dataset capturing Activities of Daily Living from multiple synchronized modalities

Siddharth Ravi[1][0000-0002-2301-569X], Pau Climent-Perez[1], Théo Morales[2], Carlo Huesca-Spairani[1], Kooshan Hashemifard[1], Francisco Florez-Revuelta[1]

[1]University of Alicante [2]Trinity College Dublin

{siddharth.ravi, pau.climent, k.hashemifard, francisco.florez}@ua.es, moralest@tcd.ie, chs17@alu.ua.es

Abstract

We introduce ODIN (the OmniDirectional INdoor dataset), the first large-scale multi-modal dataset aimed at spurring research using top-view omnidirectional cameras in challenges related to human behaviour understanding. Recorded in real-life indoor environments with varying levels of occlusion, the dataset contains images of participants performing various activities of daily living. Along with omnidirectional images, additional synchronized modalities of data are provided. These include (1) RGB, infrared, and depth images from multiple RGB-D cameras, (2) egocentric videos, (3) physiological signals and accelerometer readings from a smart bracelet, and (4) 3D scans of the recording environments. To the best of our knowledge, ODIN is also the first dataset to provide camera-frame 3D human pose estimates for omnidirectional images, which are obtained using our novel pipeline. The project is open sourced and available at https://odin-dataset.github.io.

thy solution to these problems. These cameras are generally unobtrusive, have a larger field of view, and can provide largely unoccluded views of the environments being monitored. However, HBU challenges such as pose estimation become all the more challenging due to the viewpoint and due to the heavy distortions introduced by the lens when compared to wide-angle lenses.

The aim of this work is to introduce a new large-scale omnidirectional dataset which contains numerous synchronized modalities. This includes images and videos from cameras of different types recording participants carrying out various activities of daily living, along with their physiological data. ODIN will support research in areas as varied as human pose estimation, activity recognition, person tracking and monitoring, scene understanding, privacy preservation, biometric monitoring, novel view synthesis, generative modelling, 3D scene reconstruction, and image registration. Through our first release, we aim to promote research on 3D human pose estimation using omnidirectional cameras. Research in this area is scarce, arguably due

Publications

Weakly Supervised Human Skin Segmentation using Guidance Attention Mechanisms

Kooshan Hashemifard, Pau Climent-Perez, Francisco Florez-Revuelta
Department of Computing Technology
University of Alicante
San Vicente Del Raspeig
{k.hashemifard, pau.climent,francisco.florez}@ua.es

Abstract

Human skin segmentation is a crucial task in computer vision and biometric systems, yet it poses several challenges such as variability in skin color, pose, and illumination. This paper presents a robust data-driven skin segmentation method for a single image that addresses these challenges through the integration of contextual information and efficient network design. In addition to robustness and accuracy, the integration into real-time systems requires a careful balance between computational power, speed, and performance. The proposed method incorporates two attention modules, Body Attention and Skin Attention, that utilize contextual information to improve segmentation results. These modules draw attention to the desired areas, focusing on the body boundaries and skin pixels, respectively. Additionally, an efficient network architecture is employed in the encoder part to minimize computational power while retaining high performance. To handle the issue of noisy labels in skin datasets, the proposed method uses a weakly supervised training strategy, relying on the Skin Attention module. The results of this study demonstrate that the proposed method is comparable to, or outperforms, state-of-the-art methods on benchmark datasets.

Keywords Skin segmentation · Attention mechanism · Deep neural networks

Future Work

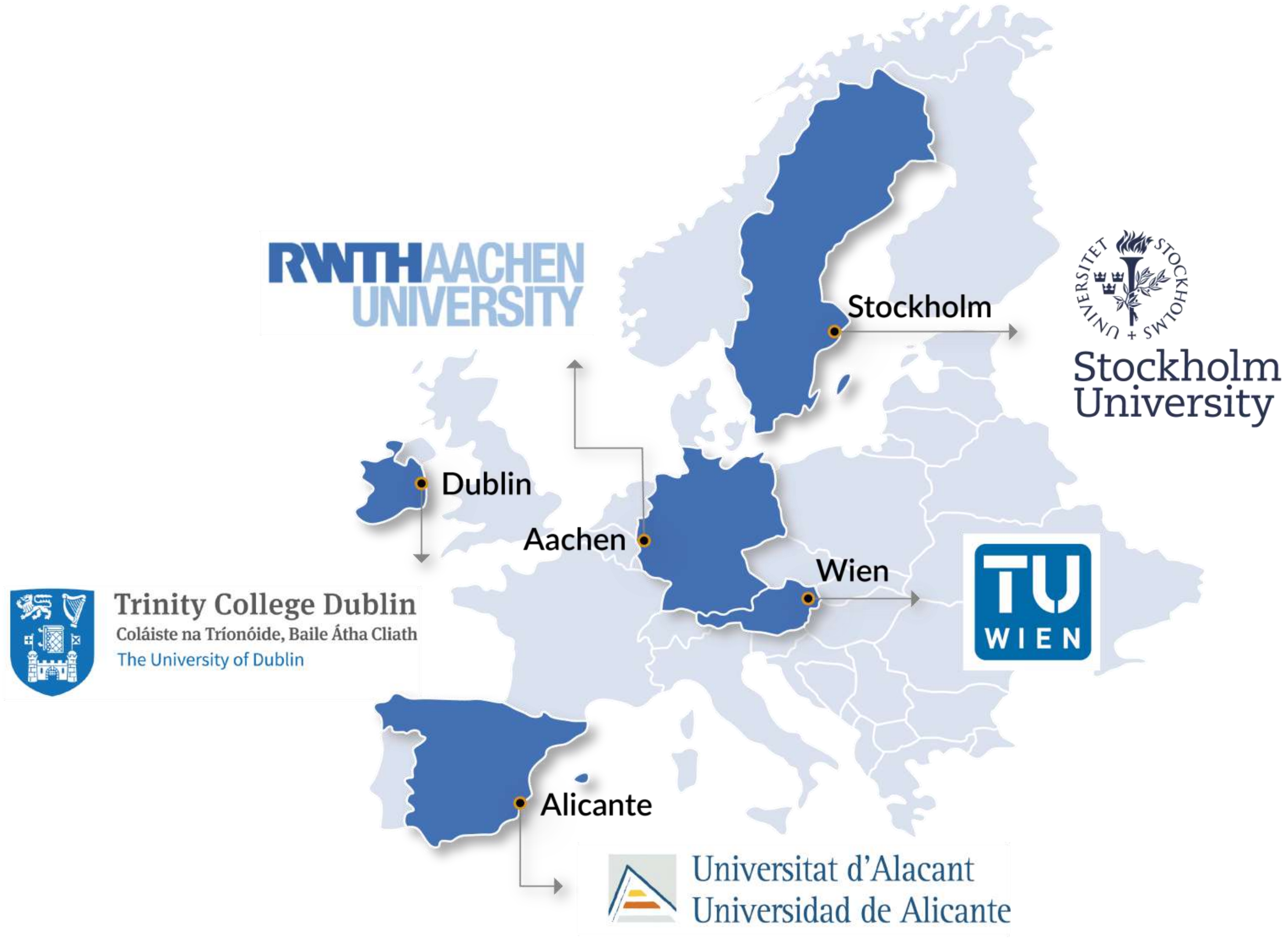
Next Variable: Activities of Daily Living in the context of AAL

- Healthcare monitoring
- Create life logging
- Long-term behavior analysis
- Remembrance and therapy adherence

- Pose and activity on Omni dataset

THIS IS THE END...

THE doors



Project Coordinator

Thank you!

Kooshan Hashemifard

University of Alicante

k.hashemifard@ua.es