HEALTH INFORMATION SYSTEM WITH PRESSURE SENSING TOWARDS A BETTER CARE PLAN FOR PRESSURE ULCERS

Undergraduate graduation project report submitted in partial fulfillment of
the requirements for the
Degree of Bachelor of Science of Engineering
in

The Department of Electronic & Telecommunication Engineering
University of Moratuwa.

Supervisors:
Dr Pujitha Silva

Group Members:
T M Piyadigama (160490F)
W A H D Perera (160480B)
T W H Tribuwan (160637N)
K E B I Edirisinghe (160140J)

August, 2021

Approval of the Department of Electronic & Telecommunication Engineering

. . . .
Head, Department of Electronic &
Telecommunication Engineering

This is to certify that I/we have read this project and that in my/our opinion it is fully adequate, in scope and quality, as an Undergraduate Graduation Project.

Supervisor: Dr Pujitha Silva

Signature: . . . .

Date: . . . .

Declaration

This declaration is made on August 07, 2021.

Declaration by Project Group
We declare that the dissertation entitled Project Name and the work presented in it are our own. We confirm that:

- this work was done wholly or mainly in candidature for a B.Sc. Engineering degree at this university,
- where any part of this dissertation has previously been submitted for a degree or any other qualification at this university or any other institute, has been clearly stated,
- where we have consulted the published work of others, is always clearly attributed,
- where we have quoted from the work of others, the source is always given,
- with the exception of such quotations, this dissertation is entirely our own work,
- we have acknowledged all main sources of help,
- parts of this dissertation have been published. (see List of Publications)

. . . .
Date

. . . .
T M Piyadigama (160490F)

. . . .
W A H D Perera (160480B)

. . . .
T W H Tribuwan (160637N)

. . . .
K E B I Edirisinghe (160140J)

Declaration by Supervisor

I/We have supervised and accepted this dissertation for the submission of the degree.

. . . .
Dr Pujitha Silva Date

Abstract

REALTIME MULTI-OBJECT TRACKING AND PIXELWISE SEGMENTATION

Group Members: T M Piyadigama, W A H D Perera, T W H Tribuwan,
K E B I Edirisinghe

Supervisors: Dr Pujitha Silva

Keywords: Vision, Perception, Detection, Tracking, Panoptic Segmentation, Siamese Network, Conditional Random Field, Recurrent Neural Network, Autonomous Systems.

Bleeding-edge technological pursuits ranging from self-guided robots at the research stage to mass scale industrial applications such as augmented reality, intelligent security systems and self-driving vehicles heavily rely on perception through vision. Vision based perception of the environment in autonomous systems extensively use object detection, segmentation and tracking as fundamental components. Despite the recent advancements in deep learning-based object detection on monocular images, several highly publicized accidents involving self-driving vehicles and critical failures in monitoring systems highlight the need for significant further improvement on real-time tracking systems in practice. We identify two such key areas with room for improvement and introduce two separate novel frameworks to tackle each problem.

We observe that trackers often perform poorly in object dense situations where occlusions and crossovers are prevalent. We identify that in order to perform better in these scenarios both appearance and motion information should be incorporated. Siamese networks have recently become highly successful at appearance based single object tracking while Recurrent Neural Networks (RNNs) have started dominating motion-based tracking. Our work focuses on combining Siamese networks and RNNs to exploit both (temporally varying) appearance and motion information to build a robust framework that can also operate in real-time. We further explore heuristics-based constraints for tracking in the Birds Eye View Space for efficiently exploiting 3D information.

Our segmentation approach is based on one of the most overwhelming problems in current vision community that has full scale perception on the image, known as panoptic segmentation where pixel level identification of the entire image is done with both semantic and instance information thus integrating object classes (thing classes having countable instance segmentation) and back-ground classes (stuff, amorphous) in a single frame. We tackle the panoptic segmentation problem with a conditional random field (CRF) model. At each pixel, the semantic label and the instance label should be compatible and spatial and color consistency of the labeling has to be preserved (similar looking neighboring pixels should have the same semantic label and the instance label). To tackle this problem, we propose a fully differentiable model named Bipartite CRF (BCRF) which can be included as a trainable first class citizen in a deep network.

Dedication

To our families, friends, supervisors, and all others that supported us in this work.

Acknowledgements

We express our sincere gratitude to our supervisors, Dr Pujitha Silva for their endless guidance, support, and commitment towards the success of this project. We would also like to thank FiveAI, UK for providing resources and support in conducting certain experiments for this project.

Contents

List of Figures

List of Tables

Acronyms and Abbreviations

RNN - Recurrent Neural Network

BEV - Bird's Eye View

CRF - Conditional Random Field

BCRF - Bipartite Conditional Random Field

LSTM - Long Short Term Memory

CNN - Convolutional Neural Network

MOT - Multi-Object Tracking

FCN - Fully Convolutional Neural Network

RCNN - Region Convolutional Neural Network

RoI - Region of Interest

IoU - Intersection over Union

MRF - Markov Random Field

AP - Average Precision

MOTA - Multiple Object Tracking Accuracy

MOTP - Multiple Object Tracking Precision

MT - Mostly Tracked

ML - Mostly Lost

SGD - Stoachastic Gradient Descent

PQ - Panoptic Quality

SQ - Segmentation Quality

RQ - Recognition Quality

1 Introduction

Pressure ulcers (which are also known as decubitus ulcers or bed sores) are a major problem in health care due to high prevalence and high cost of treatment. Some ulcers do not heal for decades. If not properly managed, pressure ulcers may cause complications such as septicemia or even death. Early prevention of pressure ulcers is beneficial over curing.

1.1 Literature Review

Prolonged external pressure into bony areas of body causes pressure ulcers in bed-ridden patients. Prevailing patho-physiological understanding of pressure ulceration is very incomplete. Several existing theories suggest that reduction of oxygen supply (under external pressure) to skin tissues causes cell death through an ischaemia-reperfusion cycle, which results in pressure ulcer formation. Another theory suggests that internal pressure on muscle tissues by bones causes ulceration. None of these theories are empirically verified. Ischaemia reperfusion models describe ulceration as a phenomenon starts at the skin and spread deep where as the last theory describe it as a phenomenon starts at muscle tissues closer the bones and spreads in the opposite direction towards bones.

Reswick and Rogers studied effect of pressure and time on cell death in 1976 modelling ischaemia oriented theory. There is no considerable improvement since then other than a few papers suggesting slight modifications. Few experiments that were ever conducted on reperfusion theory also provide inconclusive values. There is no satisfactory empirical research on internal pressure theory, although Deep Tissue Injury (DTI), a recently defined category of pressure ulcers, is specifically and widely believed to be a result of internal pressure from bones.

Exact bio-mechanical impact of pressure on human body (skin and muscles) is not yet known given that only a few unsatisfactory bio-mechanical models from early research based on animal testing and qualitative speculations without quantitative data are available. There is no sufficient data on the mechanism through which external pressure causes an ischemia-reperfusion cycle. Furthermore, there is no quantitative empirical evidence for the effect of other proposed bio-mechanical factors such as shear, friction and moisture also.

Pressure ulcers are usually located in specific sites of the body such as back of head, shoulders, buttocks, knees, elbows, hips and heels. Pressure ulcers occur in four stages according to NPUAP staging system and there are specific guideline for treatment and each stage. There is no empirical data available on how these factors affect different parts

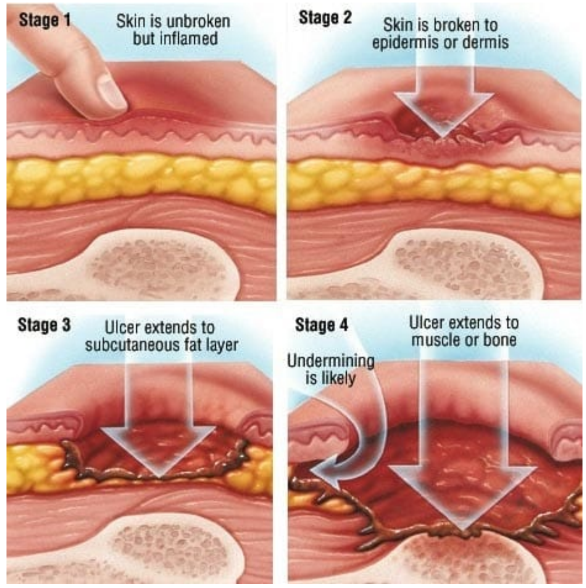

Figure 1.1: Stages I-IV of pressure ulcers according to NPUAP classifications. Image credits: https://www.nursinghomelawcenter.org/bed-sore-pictures.html

of body and different type of bodies. No empirical data is available on how the healing rates or reulceration (reulceration is not prominent as in the case of diabetic foot ulceration) was affected by the pressure.

1.1.1 Pressure Ulcer Prevention

The main pressure ulcer prevention strategy which is strongly recommended by health care authorities is frequent patient repositioning. This strategy was popularized after the end of World War II by Ludwig Guttmann, based on face validity. Caretakers should turn the patients into a different sleeping posture for every 2 h (This duration was Guttmann's ad hoc recommendation).

The absence of high quality research evidence supporting the efficacy of repositioning was discussed in a Cochrane systematic review published in 2014 (and updated in 2020). Research does not show significant advantage of 2 h repositioning over alternative time periods or no-repositioning. Currently available data are low certain and not sufficiently reliable to provide a conclusion. Existing studies do not consider bio-mechanical facts explicitly. Recently some researchers castigated the repositioning strategy for side effects such as disturbance to sleeping patterns, negative impact on dementia patients and back pain of caretakers. Recently NICE guidelines increased the time period from 2 h to 6 h for normal patients and 4 h for patients in high-risk category. Standard guidlines do not recommend to alter repositioning plans according to existing ulcers and there are no research ever conducted in that area. The efficacy of other proposed prevention strategies

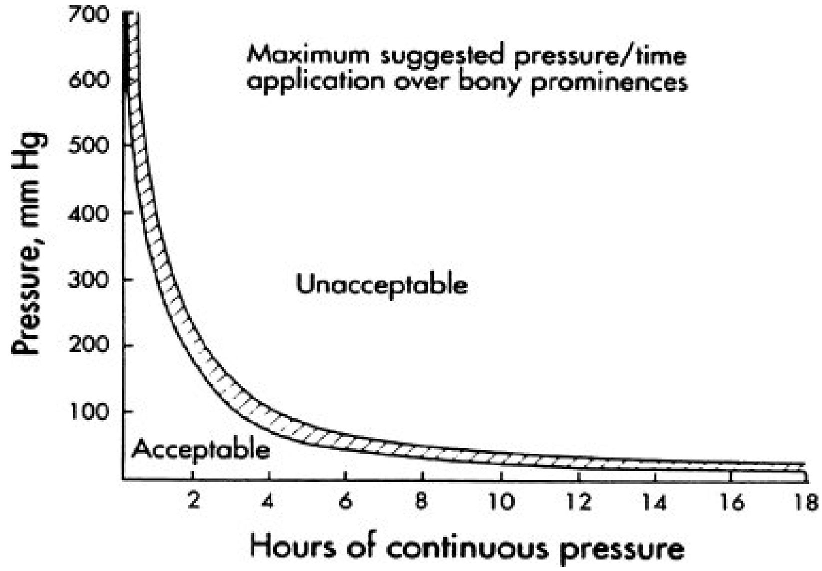

Figure 1.2: Reswick and Rogers' pressure time cell threshold graph.

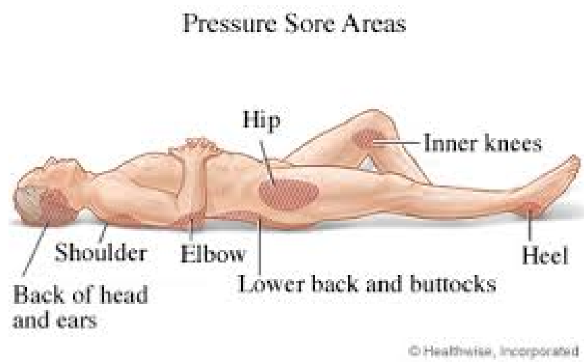

Figure 1.3: Common sites where ulceration occurs.

including the use of pressure redistribution surfaces also are not supported with research, contrary to the availability of wide range of products in the market.

1.1.2 Personal Risk Assessment and Documentation

There are several patient risk assessment indicators including Braden and Waterlow scales. The importance of a systematic scale for risk assessment is often emphasized over using clinical judgement alone. Clinical evidence on the efficacy of these tools is still insufficient and uncertain. Proper documentation is of crucial importance in modern health care. There are several paper-based or electronic documentation systems for pressure ulcers. According to studies, the purpose of existing documentation systems is not met with ulcer prevention and care. The patient repositioning plans are rarely documented. There are no records of bio-mechanical data. Existing electronic documentation systems are desktop applications that store records inside a single end-user device. The recent advancement of mobile, web, IOT and cloud technologies are not yet employed for pressure ulcer documentation.

1.2 Requirement of A Health Information System

Ajami and Khalegi discussed the importance of a wireless sensor network for pressure monitoring. There is need for a sophisticated Health Information System (HIS) that supports not only remote pressure monitoring and electronic documentation but also important utilities to optimize care plan such as posture detection, ulceration point (the specific areas of body which are more prone to ulceration) detection, pressure/risk estimation, repositioning schedule calculation and carer notification. Although there are several implementations addressing subsets of above tasks already, constructing a health information system in a holistic point of view is a novel concern. Such information system should network patients, caretakers, guardians and doctors together and generate, collect, store, analyse and interpret data for better care planning. Considering the fact that pressure ulcer is a grey area of medical research the information system should work as tool to investigate biomechanical and clinical details of bed-ridden patients. Another requirement is to provide support for semi-automation of patient care through care plan optimization like reposition schedule calculation and carer notification system. Wide availability of mobile technology paves a way for a flexible and more sophisticated reporting system. Integrating low cost pressure sensing equipment to the information system provides opportunity to monitor and collect data for long periods and to widen the scope of research including the majority of ulcer-prone patients that reside in household settings. Reduction of cost will facilitate research into pressure ulcers in developing countries.

1.3 Previous Research

There are several research into pressure monitoring equipment for bed-ridden patients. Most of those research are based on pressure measurement posture detection and pressure image segmentation. Some research considered automatic patient repositioning by actuators. The contribution of the researchers of University of Dallas has considered a wide range of aspects that are related to pressure ulceration phenomenon. These researchers paid attention to the biomechanics of pressure ulceration. But this is prior to 2014, the year several systematic reviews were published questioning the efficacy of prevention methods. Therefore the limitations to their study is not apparent in their original papers. They purposively conflate data related to different theories, scenarios and settings to achieve final results. Therefore their results are considerably depended on ad-hoc assumptions and speculations from indirect data. In this research we adopted some of their results for our purpose as the best solution available carefully evaluating the limitations.

2 Methodology

An information system architecture that supports care planning of pressure ulcers requires certain basic functionalities such as

- capturing bio-mechanical data of the body
- Analyzing those data
- collecting risk assessment data
- providing platform to report and document ulcers
- networking related people
- planning schedules.

Therefore our information system architecture is supported by pressure sensing mats and mobile apps. Some standard risk assessment scales, ulcer documentation formats are added to the system with appropriate modifications. Additionally scalability, flexibility and cost-effectiveness are two other important characteristics of such system. Our initial scope was to build a pressure sensing mattress system that is capable of recommending optimal repositioning strategies based on bio-mechanical data. As there are no proper evaluation criteria to assess pressure ulcer prevention and the existing bio-mechanical and pathological research in pressure ulcers are inconclusive, we constructed an information system that provide a platform to investigate pressure ulceration phenomenon while providing a tool for care planning by digitizing processes currently done in paper or not done in any systematic way. Existing theories can be used in our system to improve care planning within their limitation.

2.1 Components and Functionality

There are three main components of our solution.

- Information System (Server and Backend)
- Mobile App
- Pressure sensing matress

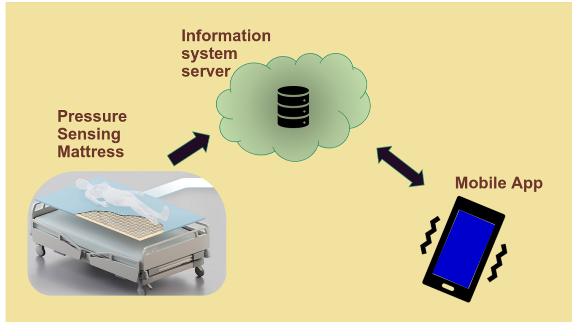

Figure 2.1: Component diagram

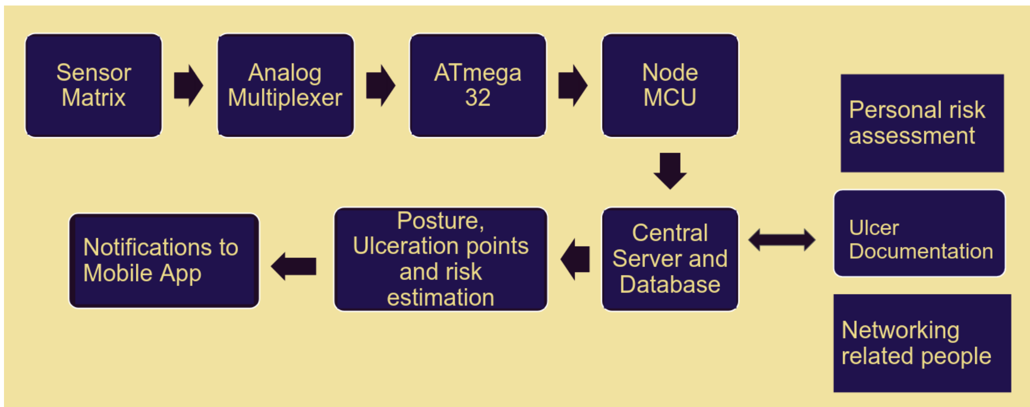

Figure 2.2: Functional block diagram

The information system provides basic components of authentication, and data storage for pressure data, personal risk assessment data and ulcer documentation. It is consist of another supplementary subcomponent for machine learning models. The information system provides a RESTful API for mobile app clients and pressure sensing mats. App and pressure sensing mat can send and retrieve relavent information from the information system. The mobile app provides user interface for patients/guardians, caretakers and doctors to interacts with the system.

The pressure mat consist of a sensor panel developed by a substrate of piezo-resistive material Velostat®. The sensor readings are processed one cell by one in the ATMega32® microcontroller and send to the information system using a NodeMCU/ESP8266® via Wifi and internet. The information system is capable of integrating other available commercial pressure sensing mattresses without any change of its structure.

In the central server of the information system these pressure data will be filtered and stored. The sleeping postures and ulceration points are identified by these data and pressure at these points is saved in a seperate table using Neural Network Models.

There is a notificaation system that sends notifications to the caretakers of patients instructing the repostitioning plan.

2.2 Information system back-end

Information system backend is written in Python using the enterprise level web fullstack designing framework Django® and hosted in Heroku® cloud platform. As the database management system we choose Postgresql which is a SQL based relational database management system. All the static media files are stored in a Cloudinary S3 bucket®. APIs are created from django-rest-framework library and Firebase® is used to communicate with mobile apps with push notifications.

The web application considered on Django apps (submodules) for each main functionality.

1. Authentication and User Profiles
2. Social connection handling
3. Pressure data
4. Personal Risk Analysis
5. Ulcer Documentaion

Neural network models are build and trained using Tensorlfow® and Keras® libraries and hosted in Heroku® using popular python backend microframework Flask®.

2.2.1 Authentication and Authorization

There are user accounts to authenticate the users and there are three groups as doctors, caretakers and patients. These roles and accounts are used to authorize access to particular components. Only users have write or update permission to their personal information, care takers can update there risk assessment data while doctors can update ulcer reporting documentation as well as risk assessment data. Even latter data only accessible to caretakers or doctors who are assigned to relevant patients. Token authentication is used to authenticate access.

To create and account a user is requested to add his username and password and there after he has to use that username and password to log in. Users can update their profile with basic details and profile photos.

2.2.2 Social Networking

All doctors, caretakers, patients can be see each other in search lists. The connection between the users are established via request and confirm mechanism. There are send, show, accept, reject, delete functionalities for a request. Doctors and caretakers can only access data of a patient only if they has been connected to the particular patient. Users can remove others from there connection list.

2.2.3 Pressure data

Pressure data sent from pressure mats are stored in the database via the central server. These data are further analysed with Neural Network Models to find ulceration points. Pressure data is stored in the format **lx, ly, x, y, p, n** format.

Here,

lx : Number of cells over x axis of the mat

ly : Number of cells over y axis of the mat

x : x coordinate of the current cell

y : y coordinate of the current cell

p : Pressure at the (x,y) cell

n : frame number (Reading complete mat is a one frame)

This format supports to send cells one by one therefore we can capture even a partial reading. This format do not restrict the resolution to a particular value we decide so changing lx and ly of the request any available pressure sensing mattress can be integrated without any structural change of the system.

2.2.4 Machine Learning

There are two machine learning models to analyze pressure data. One is to identify posture and the other is to identify ulceration points. Since the ulceration occurs in these particular sites it is important to identify pressure at those locations. To locate this points on the pressure mat and to identify repositioning we should find postures of the patient from pressure data. We used a dataset by university of Dallas to train the neural networks and we used data preprocessing and augmentations to improve the model. There are 13 people in 18 postures (5 major postures supine, left yearning, right yerning, left fetal, right fetal and there slight varaiations with rolling angle and using pillows as wedges.) There are collection of pressure distribution measured by a commercial pressure measuring mattress for 2 mins which is roughly 120 frames for each. The resoulution is 64×32.

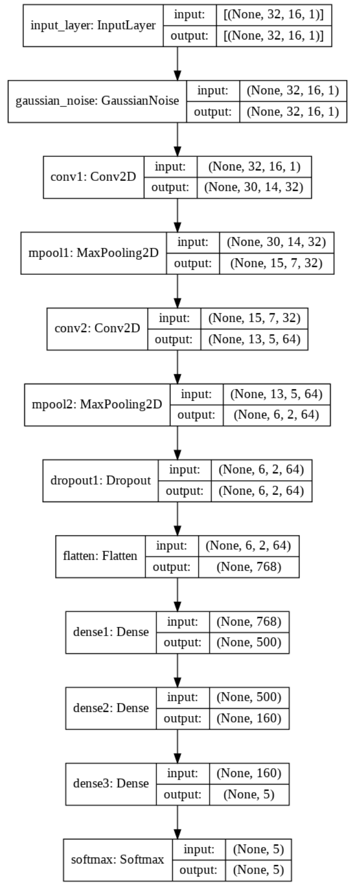

Figure 2.3: Neural Network architecture of the posture detection model. When pressure image is given the model identifies the sleeping posture

Posture Detection Model

Posture detection model is a sequential model with several convolution and pooling layers before final dense layers. When the input pressure image is given the model outputs the corresponding name of the sleeping posture.

Validation The dataset was divided into a training and holdoutset such that data from 9 persons for train and data from 4 persons for holdout.

preprocessing The pressure images are resized to 32×16, Gaussian noise of variation of 0.08 is added and finally Gaussian filter of variation of 0.5 is applied. This adding extra noise is supposed to regularize the neural network to work in more realistic environments with low cost pressure mattresses.

Data Augmentation Pressure images are rotates in random angles between by -150 to -150 and Gaussian Noise of variance of 0.1 is added.

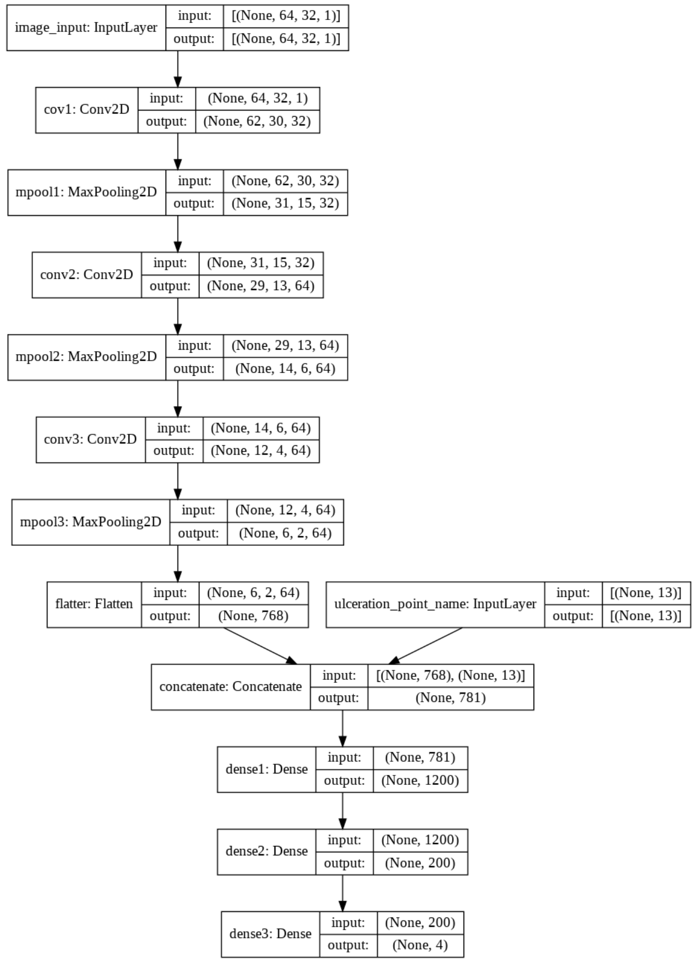

Figure 2.4: Neural Network architecture of the model for ulceration point detection. When the pressure image and the ulceration points we consider is given the model provides parameter of the bounding box for that particular site.

The 5 labels for supine, left yearning, right yerning, left fetal and right fetal are onehot encoded.

Neural network provided 92.45% holdout set accuracy.

Ulceration Point Detection Model

The same datset was used to train the neural network model for ulceration point detection. We manually created bounding boxes for ulceration points using the annotator tool Labelbox®. Then we preprocessed images likewise in the previous model. The four paramters (two coordinates of the upper left corner of the bounding box, height and width) were used to train the model with a mean squared error loss function. There are two inputs to the model. The pressure image and the name of the ulceration point we consider (onehot encoded). Then the model outputs four parameters for the bounding box.

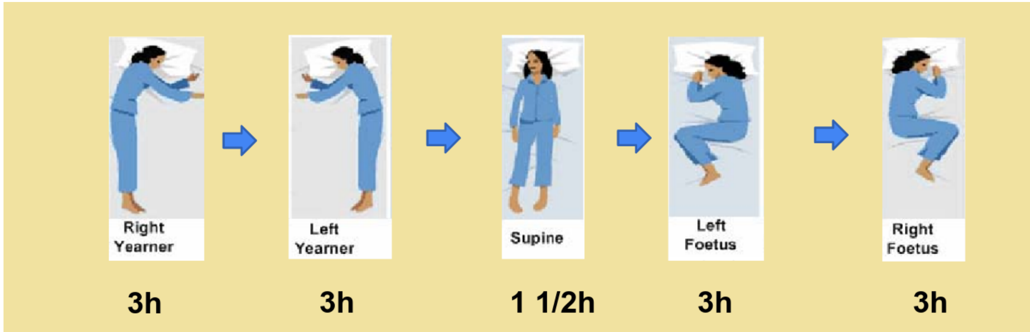

Figure 2.5: Schedule for reposition.

2.2.5 *Scheduling*

Usually 2h recommendation period for any posture is used an there is no particular order of posture order. However the researchers of University of Dallas tried to find a reposition schedule based on pressure distribution. Unfortunately there risk assessment metric is based on data from closely related research for slightly different problems and their final result is depend on ad-hoc assumptions they used.

In summary if we put the outline of their research perspectively it states that the supine posture is more risky as the both sides of the body is subjected to pressure. Although a side of body subjected to more pressure in left or right posture that is a complete relieving phase for the other side of the body. Although we hessitate about the validity of their arbitrary risk metric and ad-hoc assumption we decided to use their result and recommend a repositioning plan as follows.

1. Right Yearning - 3 h
2. Left Yearning - 3 h
3. Supine - 1.5 h
4. Left Fetal - 3 h
5. Right Fetal - 3 h

The left and right postures should be alternatively applied but we do not distinguish between yearning and fetal. As this intervals are below the range of NICE guidelines it could be justified to use these intervals.

2.2.6 *Personal Risk Assessment*

We considered two existing personal risk assessment scales Braden scale and Waterloo scale. The information system captures data relevent to both scales and calculate corresponding metrics.

The personal risk assessment forms contains following data and expected to be filled by a health care professional (a doctor or a nurse).

Assessed By : The doctor or the caretaker (nurse) assessed personal risk

Gender : Male/Female

Age : Age of the patient

Weight : Weight of the patient (kg)

Height : Height of the patinet (cm)

These details should be filled as 1,2,3,4 according to the Braden scale guideline.

Sensory perception : Ability to respond meaningfully to pressure-related discomfort

Moisture : Degree to which skin is exposed to moisture

Activity : Degree of physical activity

Mobility : Ability to change and control body position

Nutrition : Usual food intake pattern

Friction and Shear

Explicit definition of 1,2,3,4 levels for each category is given in the Braden scale guideline (which is showed by an information box in the mobile app.)

According to the total score the patients are classified into four risk categories.

Severe risk	≤ 9
High risk	10 - 12
Moderate risk	13 - 14
Mild risk	15 - 18

These are some other important risk factors (Yes/No binary options).

- Diabetes mellitus
- Peripheral vascular disease
- Cerebral vascular accident
- Hypotension

- Hypoalbuminemia
- Incontinence
- Venus thrombosis

2.2.7 Ulcer documentation

Documenting existing ulcers is an important concern. Treatments are based on proper documentation. This includes basic details related to the wound, surrounding skin and conditions of the patient. We adopted basic components from NPUAP (National Pressure Ulcer Advisory Panel) guidelines and SOS (State of Oklahoma) toolkit to prepare our documentation patteren. We discussed the current state of pressure ulcer documentation with a medical practitioner in Sri Lanka and remove overcomplicating components from these two guidelines. Then we add several additional components and alter the terminology in order to make compatible with medical terminology used in Sri Lanka. Some of the components we introduce here are not currently documented in Sri Lanka.

Reported by : The doctor that reports the ulcer

ChangeAddDelete : The updated date (automatically filled)

Site : Ulceration points

Stage : Stage I,II,III,IV, DTI (Deep Tissue Injury), Unstaged (NPUAP classification)

Duration : Duration (days)

Length : Length of the ulcer (mm)

Width : Width of the ulcer (mm)

Depth : Depth of the ulcer (mm)

Margin : Regular, Irregular

Edge : Sloping, Punched out, Rollout, Everted

Edge color : Color of the edge of the ulcer

Underminings : (Yes/No)

Sinus tracts : (Yes/No)

Floor : Healthy, Granualation Tissue, Slough, Necrotic, Eschar, Epithelial (Multiple selection)

Discharge : Serous, Purulent, Serosanguineous, Other

Discharge amount : Small, Medium, Heavy

Surrounding skin : Warm, Thickend, Hyperpigmented, Hypopignmented, Gangreous, Itching, Swelling (Multiple selection)

Skin sensation : Good, Impaired

Regional lymph nodes enlarged : Yes/No

Smell : Yes/No

Pain : Yes/No

Progress : Improved, No change, Stable, Decline

Image : Image of the ulcer

2.3 Mobile app

Mobile app provides a user interface for basic funcionalities of the system. This includes,

- Login
- Profile update
- Search other users
- Handle social connections
- Register mattress
- Personal Risk Assessment
- Ulcer documentation
- Notification

Notifications are send 5 mins prior to the reposition and another at the moment reposition is planned. Next posture and the period in that posture is given with the notification. If the patient is not turned at the specified time then another notifications are send three times with a 5 min interval.

Figure 2.6: Universal testing machine

2.4 Pressure Mat

There are two difference methods to create a pressure mat. The first method is to combine large number of sensors and the second method is to devevlop a single substrate of pressure sensing material into a sensor panel. First approach is manufacturably complex. Therefore we selected second approach. Velostat® is a low cost piezo-resistive material that is used for similiar applications. Selection of piezoresistive material over piezocapacitive material reduce complexity of the senosor interfacing. Resistance can be measured constructing a voltage divider.

2.4.1 Calibration of the Material

2.4.2 Preparing Mat

Velostat sheet was sandwiched by two Neoprene sheets each one contains set of parallel rows or columns of copper tapes. Each column is attached to the output channels of an analog multiplexer and each rows are attached to the input channels of two multiplexers that works as a one multiplexer in combine.

Columns are powered one by one using the analog multiplexer and the voltage is measured over a voltage divider choosing each row by other multiplexers.

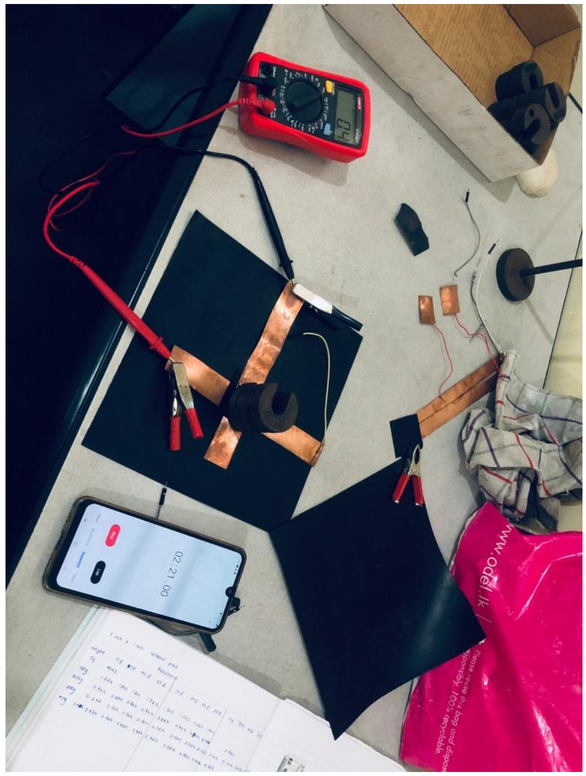

Figure 2.7: Measuring with deadweight

Neoprene acts as an insulator to build the copper wire grid. All rows are weakly pulled down according to previous research which shows it reduce cross-talk effects.

2.4.3 Sensor reading processing and communication

The communication between ATMega32 and ESP8266 is via UART and then the WIFI server sends it to server.

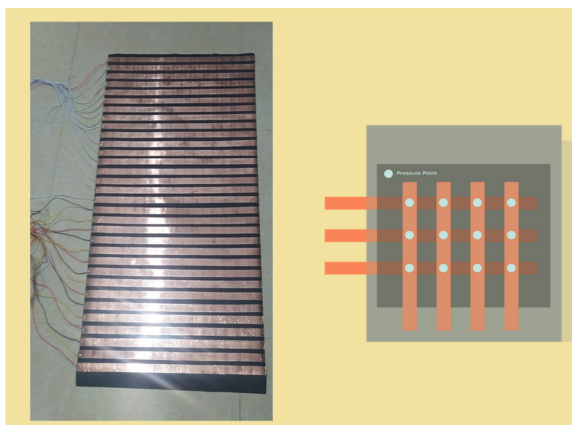

Figure 2.8: Pressure mat

3 Results

Figure 3.1: Pressure mat frames

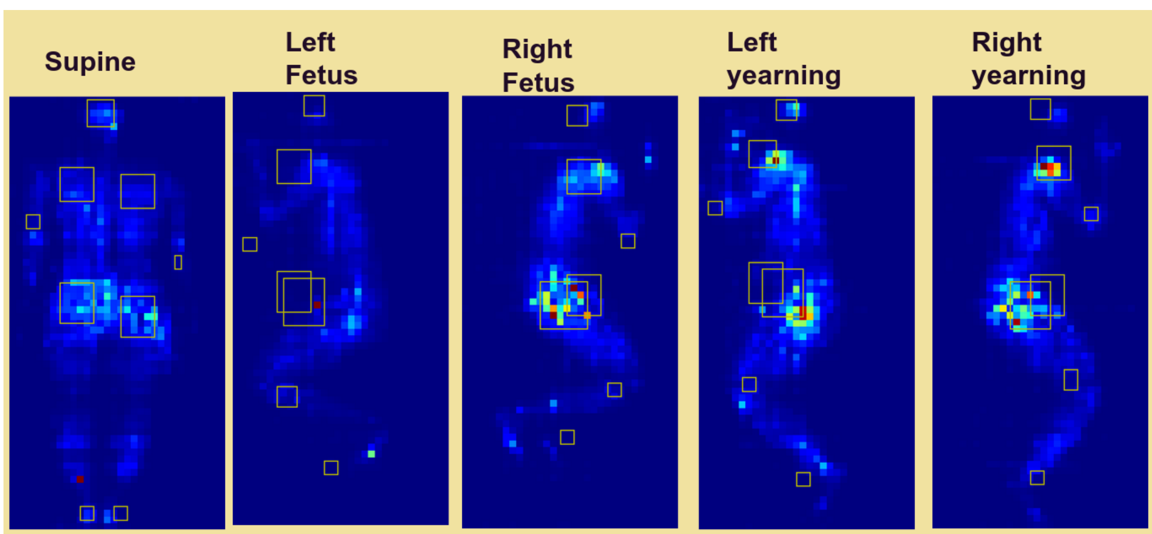

Figure 3.2: Neural Network result

4 Discussion and Conclusion

We propose two components essential for autonomous systems that interact with their surrounding environments. These are in fact two of the key computer vision problems that have been attempted for a long time.

Firstly, we present an end-to-end system capable of performing multi-object tracking by combining a range of advances in object detection and reidentification along with our novel architectures and loss functions. Further, we work on a novel step by building a separate LSTM branch to estimate the similarity feature map for the next time step of a given track. The Siamese Networks may be viewed as a two-step version of our extension, whereas this replacement with an LSTM is more of a generalized version capable of generating a better feature set. The key expectation with this addition is the overcoming of identity switches and lost tracks in the case of occlusions. Appearance features tend to change significantly during an occlusion, especially when an object undergoes rotations, and our extension overcomes this by modeling the appearance changing pattern over time.

Thereafter, we proposed a probabilistic graphical model based framework for panoptic segmentation. Our CRF model with two different kinds of random variable, named Bipartite CRF or BCRF, is capable of optimally combining the predictions from a semantic segmentation model and an instance segmentation model to obtain a good panoptic segmentation. We use different energy functions in our BCRF to encourage the spatial, appearance, and instance-to-semantic consistency of the panoptic segmentation. An iterative mean field algorithm was then used to find the panoptic labeling that approximately maximizes the conditional probability of the labeling given the image. We further showed that the proposed BCRF framework can be used as an embedded module within a deep neural network to obtain superior results in panoptic segmentation.

4.1 Principles, Relationships and Generalizations inferred from results

As depicted in the results section, our tracker has shown improvements basically in relation to MOT evaluation metrics. The improvements presented based on the KITTI dataset (which has 9 separate classes) shows how our system has generalized multi class tracking without the need for training separate computationally expensive re-identification networks. MOT16 contains data belonging to the pedestrian class only but the movement of objects in this class is subjugated to more occlusions and random movements compared to the KITTI dataset. The improvement of MOTA over MOT16 dataset indicates signs that our system handles occlusions better. It is also evident not only through the dataset statistics but also through the visual online videos that our system has less number of lost tracks in the middle of a certain scenario.

In panoptic segmentation, the results depict the principle analysis that bipartite conditional random fields propose an improved labeling in both semantic as well as instance domains where initial unary potentials for semantic and instance identities are taken from unary classifiers that are state of the art systems at present. The results also show that the cross potential component of the aggregated energy function that is being minimized during an inference has effects beyond rest of the energy function with both semantic component and instance component separately. The improvements observed in the Panoptic Quality are also visually consistent with intuition that stray patches of the final output have mostly been removed and the edges of objects have been smoothened. The final output when split and analyzed semantically and instance wise; the qualitative results present the consistency and clarity in comparison to the unary classifier outputs.

4.2 Problems and Exceptions to the Generalizations

The results show that MOTP of our tracker is considerably low in MOT16 dataset in comparison to other systems. This indicates that the LSTM network is unable to handle rapid variations of the bounding box parameters. This is to be expected as the bounding box variations in datasets such as MOT16 is extremely chaotic in cases where the pedestrian is rotating while walking and moving in general. This is also due to the morphological changes of the moving body specifically a bounding box is not an ideal interpretation of the object. The hand gesture changes are also changing the bounding box co-ordinates of the object considered. However this complication does not arise for the cases where automobiles are considered. It was also observed that system has higher performance in time domain when automobile motion is considered.

The system implemented for panoptic segmentation through the aggregation of two separate heads built for semantic and instance segmentation having state of the art accuracy builds up a compatibility matrix that compares the class wise cross compatibility of the instance and semantic classes. This learns the entry matrix elements from the dataset. However if the dataset is biased for say person class (As in the case of Pascal VOC); there is a tendency of having arbitrarily high compatibility which is dataset dependent. This can be avoided by using large datasets which are robust however that training task requires considerable computational resources.

4.3 Agreements/Disagreements with previously published work

The results agree with recently published systems such as Deep SORT [1]. It is expected that as ML decreases when MOTA increases as it reduces the number of false negatives considerably. This correlation is depicted in our results. However the experiments that had been run basing the data association LSTM network did not turn successful as presented in [2]. However in [2] it describes as a network not promised to have high accuracy but possesses higher frame rate in comparison to the accuracy. As a result, the lack of data association capability and the retarded smoothness in convergence could be expected when single module is isolated from the aforementioned network and tried to train starting from Xavier initialization.

We were able to replicate the recent most state of the art systems to obtain the unary classifications on image segmentation. The approach followed by our system agrees with the work published by authors in [3] for refining the output of a single head semantic segmentation network using conditional random fields. Our system was integrated on top of a state of the art system presented in [76]. We used the loss function presented in [4] for training.

References

[1] N. Wojke, A. Bewley, and D. Paulus, "Simple online and realtime tracking with a deep association metric," in *IEEE International Conference on Image Processing (ICIP)*, 2017.

[2] A. Milan, L. Leal-Taixé, I. D. Reid, S. Roth, and K. Schindler, "MOT16: A benchmark for multi-object tracking," *CoRR*, vol. abs/1603.00831, 2016. [Online]. Available: http://arxiv.org/abs/1603.00831

[3] S. Zheng, S. Jayasumana, B. Romera-Paredes, V. Vineet, Z. Su, D. Du, C. Huang, and P. Torr, "Conditional Random Fields as Recurrent Neural Networks," in *ICCV*, 2015.

[4] A. Arnab and P. H. Torr, "Pixelwise Instance Segmentation with a Dynamically Instantiated Network," in *CVPR*, 2017.

[5] J. F. Arsalan Mousavian, Dragomir Anguelov, "3D Bounding Box Estimation Using Deep Learning and Geometry," in *Conference on Computer Vision and Pattern Recognition (CVPR)*, 2017.

Appendix I

Ability to track multiple objects in BEV space and the possible usage of heuristics in BEV space is explained here as an extension of the single image based tracking method presented earlier.

Extensibility to 3D tracking

Here we use the concept that objects cannot overlap in Birds Eye View space. An LSTM network is trained to predict the change of parameter q between consecutive frames. That is, for given $\dot{q_{t-k}}, ..., \dot{q_{t-1}}, \dot{q_t} \rightarrow \dot{q_{t+1}}$ is predicted where $\dot{q_t} = q_t - q_{t-1}$ and $q \in (C, S, \theta)$. Here;$C = C_x, C_y, C_z$ (the centre co-ordinates of the object), $S = (h, w, l)$ (object dimensions) and θ is the angle of rotation around the vertical axis. The loss function for training the parameter predictor (LSTM) is as follows.

$$LOSS_{pred}(p, \beta, \alpha, \delta, \theta) = \sum_{i=1}^{N} \beta_{class_i} \left(\left(\sum_{p \in (C,S)} \alpha_p L_{Huber, \delta_p}(p_{pred}, p_{gt}) \right) + \alpha_\theta L_\theta(\theta_{pred}, \theta_{gt})_{object=i} \right) \tag{7}$$

Here P_{pred} refers to the predicted parameter and P_{gt} refers to the ground truth parameter.

δ_p is a parameter based learnable which in turn is the quadratic-linear margin of the Huber loss function and α_p or α_θ is a regressed parameter based learnable (where in the case of α_θ, the regressed parameter is θ and α_p is similarly interpreted whereas the scope of α_p is different from that of δ_p, considering the impact on cost function) and β_{classi} is the class based learnable parameter w.r.t. the class of the i^{th} object.

Here, $p = C_x, C_y, C_z, h, w, l$, $\beta = \beta_{class} | class \in classes$, $\alpha = [[\alpha_p]_{p \in parameters}, \alpha_\theta]$ and $\delta = [\delta_p]_{p \in parameters}$.

Due to the discontinuous nature of the parameter θ at the two extreme ends of its domain $[-\pi, \pi]$, and due to the fact that $\theta = \pi$ and $\theta = -\pi$ depict the same orientation, it is not directly incorporated into the Huber loss function. It is handled separately using L_θ function [5], where $\theta_{pred}, \theta_{gt}$ are predicted and ground truth values of the parameter θ respectively.

$$L_\theta(\theta_{pred}, \theta_{gt}) = 0.5(1 - \cos(\theta_{gt} - \theta_{pred})) \tag{8}$$

Constraints as penalties

First, we introduce the hard constraint on BEV space that projections of the objects on to the x-z plane in general co-ordinates have no intersection. However, most of the research is focused on building up 3D bounding boxes of objects where the rectangular projection does not create a clear cut segmentation of the object (ex: human) on BEV space. Therefore, we minimize an additional term as follows.

$$I = \sum_{v_i, v_j \in objects_{pred}, i \neq j} (1 + \xi^2_{class_i, class_j})(v_{i_{BEV}} \cap v_{j_{BEV}}) \tag{9}$$

Where $v_{i_{BEV}}$ is the projection of the bounding box of the object v_i onto the BEV space and $\xi_{class_i, classj}$ is a learnable based on object classes under intersection which in turn forms a set $\xi_{class \times class}$ and each term is squared to ensure positivity. Therefore, the final minimization function is as follows,

$$L(p, \beta, \alpha, \delta, \theta, \{\xi\}) = LOSS_{pred}(p, \beta, \alpha, \delta, \theta) + I \tag{10}$$

However, at an optimum point $(p^*, \beta^*, \alpha^*, \delta^*, \theta^*, \{\xi\}^*)$; the loss function obeys a feature observed in Lagrange constrained optimization that; $\nabla L = 0$ where ∇ refers to the discrete derivative (this statement is intuitive only with the discrete derivative).

This implies that:

$$\nabla_{p,\theta} Loss_{pred} = -(1 + \xi^2_{class_i, class_j}) \nabla_{p,\theta} (v_{i_{BEV}} \cap v_{j_{BEV}}) \tag{11}$$

for all classes at optimum parameters p^*, θ^*. Therefore $(1 + \xi^2_{class_i, class_j})$ behaves similar to a Lagrange multiplier. This setting helps to build up a network that trains not only based on the individual performance per object but also encountering the joint effect of multiple object scenarios.

Appendix II

Mean Field Algorithm

Algorithm 1 Inference on Bipartite CRF

$Q_i(l) := \mathrm{softmax}_i(-\phi_i(l))$ and $R_i(t) := \mathrm{softmax}_i(-\psi_i(t))$ ▷ Initialization
while not converged **do**
  $Q'_i(l) \mathrel{-}= \phi_i(l)$ ▷ Update due to the first term
  $Q'_i(l) \mathrel{-}= \sum_{l' \in \mathcal{L}} \left(\mu(l, l') \sum_{j \neq i} \mathrm{Sim}_\Phi(i, j)\, Q_j(l') \right)$ ▷ Update due to the second term
  $R'_i(t) \mathrel{-}= \psi_i(t)$ ▷ Update due to the third term
  $R'_i(t) \mathrel{-}= \sum_{t' \in \mathcal{T}} \left([t \neq t'] \sum_{j \neq i} \mathrm{Sim}_\Psi(i, j)\, R_j(t') \right)$ ▷ Update due to the fourth term
  $Q'_i(l) \mathrel{-}= \sum_{t \in \mathcal{T}} \left(f(l, \mathrm{class}(t))\, R_i(t) \right)$
  $R'_i(t) \mathrel{-}= \sum_{l \in \mathcal{L}} \left(f(l, \mathrm{class}(t))\, Q_i(l) \right)$ ▷ Updates due to the fifth term
  $Q'_i(l) \mathrel{-}= \sum_{t \in \mathcal{T}} \left(f(l, \mathrm{class}(t)) \sum_{j \neq i} \mathrm{Sim}_\Omega(i, j)\, R_j(t') \right)$
  $R'_i(t) \mathrel{-}= \sum_{l \in \mathcal{L}} \left(f(l, \mathrm{class}(t)) \sum_{j \neq i} \mathrm{Sim}_\Omega(i, j)\, Q_j(l') \right)$ ▷ Updates due to the sixth term
  $Q_i(l) := \mathrm{softmax}_i \left(Q'_i(l) \right)$ and $R_i(t) := \mathrm{softmax}_i \left(R'_i(t) \right)$ ▷ Normalization
end while

Appendix III

List of Publications

- Extending Multi-Object Tracking systems to better exploit appearance and 3D information

- Bipartite Conditional Random Fields for Panoptic Segmentation