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:	Group Members:
Dr Pujitha Silva	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.
DeepSiam:SiamFC

. . . .
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 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 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 app. 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 Component Diagram

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. Postegresql an enterprise level SQL based relational database management tool is used as the database management system. All the static media files are stored in a Cloudinary S3 bucket. APIs are created from django-rest-framework platform and firebase is used to communicate with mobile apps as push notifications.

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.

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 request, show, accept, reject, delete functionalities for a request.

2.2.3 Pressure data

Pressure data sent from pressure mats are stored in the database via the 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. This format allows us to send cell readings one by one and flexible the system for any resolution. This pressure data is used to create the mat.

2.2.4 Machine Learning

There are two machine learning models to analyze pressure data. One is to identify pressure and the other is to identify ulceration points

Neural network training we used a dataset by university of Dallas and we used data preprocessing and augmentations.

2.2.5 Scheduling

There is a specific risk scale by university of Dallas and using this risk scale we could find that the order is right left shifting and half of the time for supine.

2.2.6 Risk Assessment

Braden scale and additions to the risk assessment scales

2.2.7 Ulcer documentation

Ulcer documentation is based on SOS guidelines and appropriate modifications.

2.3 Mobile app

Mobile app is developed by using React Native which is a cross platform framework (Android and IOS development framework.)

Push notifications are send to the app.

2.4 Pressure Mat

The selected material was Velostat.

2.4.1 Individual sensor calibration

Material testing. Velostat was selected. We obtained curves for Velostat.

2.4.2 Preparing Mat

Multiplexers and ATMega code was written. Node MCU was used for communications. UART and cross talk handling.

3 Results

We obtain results for our two sub-tasks on selected popular datasets. The results are reported using standard metrics commonly used to evaluate these tasks.

3.1 Multi Object Tracking Evaluation

3.1.1 Datasets and Evaluation metrics

Experiments are conducted on the MOT16 [1] and KITTI [2] tracking datasets. The MOT16 dataset contains 7 videos in its training set. The KITTI tracking dataset contains 21 videos in its training set. The Siamese Network for appearance consistency is trained completely on external data (ImageNet datasets) and there is no overlap with any of the MOT16 or KITTI data. The LSTM network is trained only with the use of bounding box locations of objects and class information for a partition of the training sets of these two datasets (the remainder is kept aside for testing purposes). Results are reported for our test partition (in the case of LSTM usage) and for the entire datasets (in cases they are not used for training).

Evaluation of our system is carried out for the entire system as well as for the study of LSTM network alone. For the case of the entire system, we consider the metrics used by the MOT benchmarks for evaluation. This includes Multiple Object Tracking Accuracy (MOTA), Multiple Object Tracking Precision (MOTP), the ratio of Mostly Tracked targets (MT), and the ratio of Mostly Lost targets (ML). In the case of the LSTM network, the Average Precision (AP) value for the predicted frames across the dataset and classes is reported.

3.1.2 Evaluation

The evaluations on the MOT16 Dataset for the end to end system are reported in Table I. Evaluations mainly focus on two aspects: improvement in accuracy with the introduction of the similarity measure to a traditional tracker using only a Kalman filter or an LSTM network and how closely related the accuracy is with state of the art multi-object trackers. Similar results on the KITTI tracking dataset are presented for our work alongside comparisons (note that few state-of-the-art works report on this dataset) in Table II. Separate evaluations for the LSTM in the case of single object tracking for individual tracklets in the KITTI dataset was carried out. An average IoU of 61.45 and AP of 0.96 at 0.5 IoU were obtained for this experiment.

Table 3.1: Comparison of our performance on MOT16 dataset with recent works

Method	Mode	MOTA↑	MOTP↑	MT↑	ML↓
Deep SORT [3]	ONLINE	61.40%	79.10%	32.80%	18.20%
SORT [4]	ONLINE	59.80%	79.60%	25.40%	22.70%
RNN LSTM [5]	ONLINE	19.00%	71.00%	05.50%	45.60%
MDP [6]	ONLINE	30.30%	71.30%	13.00%	38.40%
DMAN [7]	ONLINE	46.10%	73.80%	17.40%	42.70%
LSTM+Similarity (Ours)	ONLINE	66.70%	69.00%	39.18%	16.80%
Kalman Filter (Ours)	ONLINE	61.00%	69.00%	17.00%	17.00%

Table 3.2: Comparison of our performance on KITTI-trracking dataset with recent works

Method	Mode	MOTA↑	MOTP↑	MT↑	ML↓
Regionlets Only [8]	ONLINE	76.40%	81.50%	54.10%	9.30%
MS-CNN Only [8]	ONLINE	81.23%	85.60%	66.30%	4.60%
Regionlets MS-CNN [8]	ONLINE	82.60%	85.00%	70.50%	5.30%
SMES [9]	ONLINE	70.78%	80.38%	51.68%	7.77%
LSTM + Similarity (Ours)	ONLINE	83.58%	78.50%	48.23%	2.25%

3.1.3 Experiments to analyze the extensibility of the modules

LSTM based data association for end to end trainability: The LSTM network was trained under negative log likelihood loss. It was observed that network was not developing a significant convergence even for a fixed set of data associations which was also the key expectation. This methodology is trainable but in comparison to the results from the Hungarian algorithm, the associations are sub-optimal and have a significant potential of resulting in non-coherent results (similar to observations at the training phase) which deprecate the accuracy of the entire system henceforth. The results being inconsistent as well as non-coherent and the observation that training sessions do not converge to a feasible setting made this sub module unsuccessful in terms of performance.

Feature Predictor: The feature space is significant in every novel approach considered in the fields of vision based analysis where tracking is only a sub group of it. Along with the ability of the LSTM networks to perform well in prediction, and as model predictors used in current trackers perform linear interpolation of the feature tensors, an experiment on the ability of a trainable network to predict the feature space was conducted (generic interpolation of features is done in most of the Siamese tracking networks and they are proven to perform well in practice). During this analysis, a robust LSTM network with high hidden state size was trained on the extracted features in sequences. (Ability to compare a com-

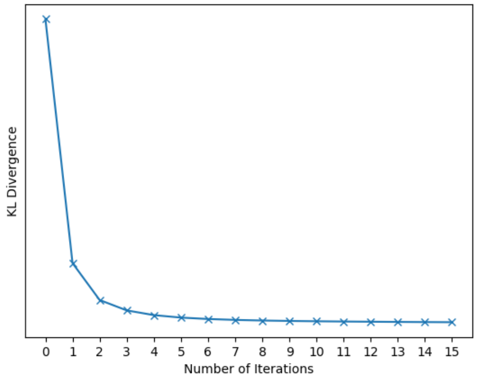

Figure 3.1: **Convergence of BCRF Inference.** Convergence of KL divergence with the number of iterations.

plete feature vector can be integrated for optical flow analysis and many further approaches if turns out to be successful). However the feature predicting network was over fitting to the dataset (custom) during the training phase. The accuracy of the tests for network validation was poor and turned out insufficient for any further analysis.

3.2 Panoptic Segmentation Evaluation

In this section, we first show the convergence of the mean field based inference algorithm for BCRF and then show the usefulness of the BCRF model by evaluating its performance on the Pascal VOC dataset and the COCO dataset.

3.2.1 Convergence of the Inference

It is difficult to provide a theoretical convergence guarantee for mean field algorithms with parallel updates [10, 11]. We therefore provide empirical evidence to show that the presented mean field inference algorithm for our BCRF with cross potentials converge under normal conditions. To this end, we estimate the KL divergence between the original joint distribution and the factorized distribution (see Eq. (**??**)), at the end of each iteration in Algorithm 1. Note that this KL divergence can be estimated up to a constant using the method described in [12]. We pick 20 random images from the Pascal VOC validation set and average the KL divergence for each iteration across these images. The resulting plot is shown in Fig. 3.1. It can be seen that the KL divergence measure, and therefore the inference algorithm, converges within a few iterations. We also note that visual results do not change after about 5 iterations.

3.2.2 Bipartite Potentials Learning

Figure 3.3 illustrates how important logits belonging to each class in the instance branch are for predicting each class in the semantic branch when the model has been fully trained.

Method	PQ	SQ	RQ
DeeperLab [16]	67.35	-	-
Ours (baseline)	70.50	88.65	78.83
Ours (CRF only)	67.72	87.62	76.48
Ours (BCRF)	71.76	89.63	79.33

Table 3.3: Comparison of results on Pascal VOC dataset. The baseline used contains DeepLab-v3 for semantic branch and Mask-RCNN for instance branch followed by combination using the simple logical method outlined in [17]. CRF only corresponds to setting the BCRF cross-potential terms to zero. BCRF is our complete network.

Our BCRF module allows the network to learn complex relationships between the semantic and instance features belonging to each class. While there is room for it to learn a simple logical relationship, the variation of learned parameters in Figure 3.3 verifies that a complex class-specific mapping has been learned by the network.

3.2.3 Results on the Pascal VOC Dataset

In this experiment we use the architecture shown in Figure **??** and CNN components similar to the ones used in [13]. More specifically, we use a ResNet-50 with an FPN as the backend, to which we attach a fully convolutional network as the semantic segmentation head and a Mask R-CNN network as the instance segmentation head.

During both training and inference we used 5 mean-field iterations for BCRF. At the output, we calculate the loss function as a summation of two components: the usual pixel-wise categorical cross entropy loss for the semantic component [14] and the loss used in [15] for the instance component. We used full-image training with batch size 1 and SGD with learning rate 0.0007 and momentum 0.99. In Table 3.3, we report the summary of the quantitative results. Table 3.4 shows the class-wise results. Qualitative results are shown in Table 3.5, where benefits of optimally combining the semantic segmentation classification and instance segmentation classification with BCRF can be seen.

3.2.4 Results on the COCO Dataset

To further evaluate the usefulness of BCRF without any efforts for end-to-end training, experiments were conducted on the COCO dataset by simply plugging in the BCRF on an existing pre-trained model. We used a combination of publicly available models of [13, 18], which produced a PQ score of 41.4% on the COCO validation set. The parameters of the BCRF were hand-tuned using a small subset of train images. Results obtained from that BCRF model without end-to-end training are listed in Table 3.6.

	PQ		**SQ**		**RQ**	
Class	W/O BCRF	BCRF	W/O BCRF	BCRF	W/O BCRF	BCRF
Background	90.8	92.33	93.39	94.69	97.22	97.51
Aeroplane	78.55	80.37	88.57	92.6	88.68	86.79
Bicycle	29.78	31.71	67.36	68.46	44.21	46.32
Bird	84.98	85.09	93.05	93.24	91.32	91.25
Boat	65.83	66.21	85.33	86.48	77.14	76.56
Bottle	67.44	64.05	92.05	90.68	73.26	70.63
Bus	82.68	82.58	94.56	95.46	87.44	86.51
Car	72.22	70.93	93.69	91.7	77.08	77.35
Cat	77.41	83.4	91.24	93.73	84.85	88.97
Chair	43.3	41.79	82.5	82.64	52.49	50.57
Cow	76.91	80.42	92.81	93.95	82.87	85.6
Diningtable	51.33	51.8	80.81	82.88	63.51	62.5
Dog	76.63	81.59	90.5	93.29	84.67	87.46
Horse	76.86	81.4	89.38	91.11	86	89.34
Motorbike	78.07	80.21	87.5	89.89	89.23	89.23
Person	76.33	77	89.75	89.73	85.05	85.81
Pottedplant	58.98	60.62	85.41	85.32	69.06	71.05
Sheep	74.29	74	93.86	93.48	79.15	79.15
Sofa	60.37	62.12	88.47	89.5	68.24	69.41
Train	78.52	80.05	88.7	90.43	88.52	88.52
Tvmonitor	79.23	79.34	92.8	92.93	85.38	85.38
Mean Value	**70.5**	**71.76**	**88.65**	**89.63**	**78.83**	**79.33**

Table 3.4: **Pascal VOC dataset.** Detailed class-wise panoptic segmentation results on the Pascal VOC validation set comparing results without BCRF vs with BCRF on a standard network.

3.2.5 BCRF learns beyond simple logical mapping

Figure 3.3 illustrates how important logits belonging to each class in the instance branch are for predicting each class in the semantic branch when the model has been fully trained. Our BCRF module allows the network to learn complex relationships between the semantic and instance features belonging to each class. While there is room for it to learn a simple logical relationship, the variation of learned parameters in Figure 3.3 verifies that a complex class-specific mapping has been learned by the network.

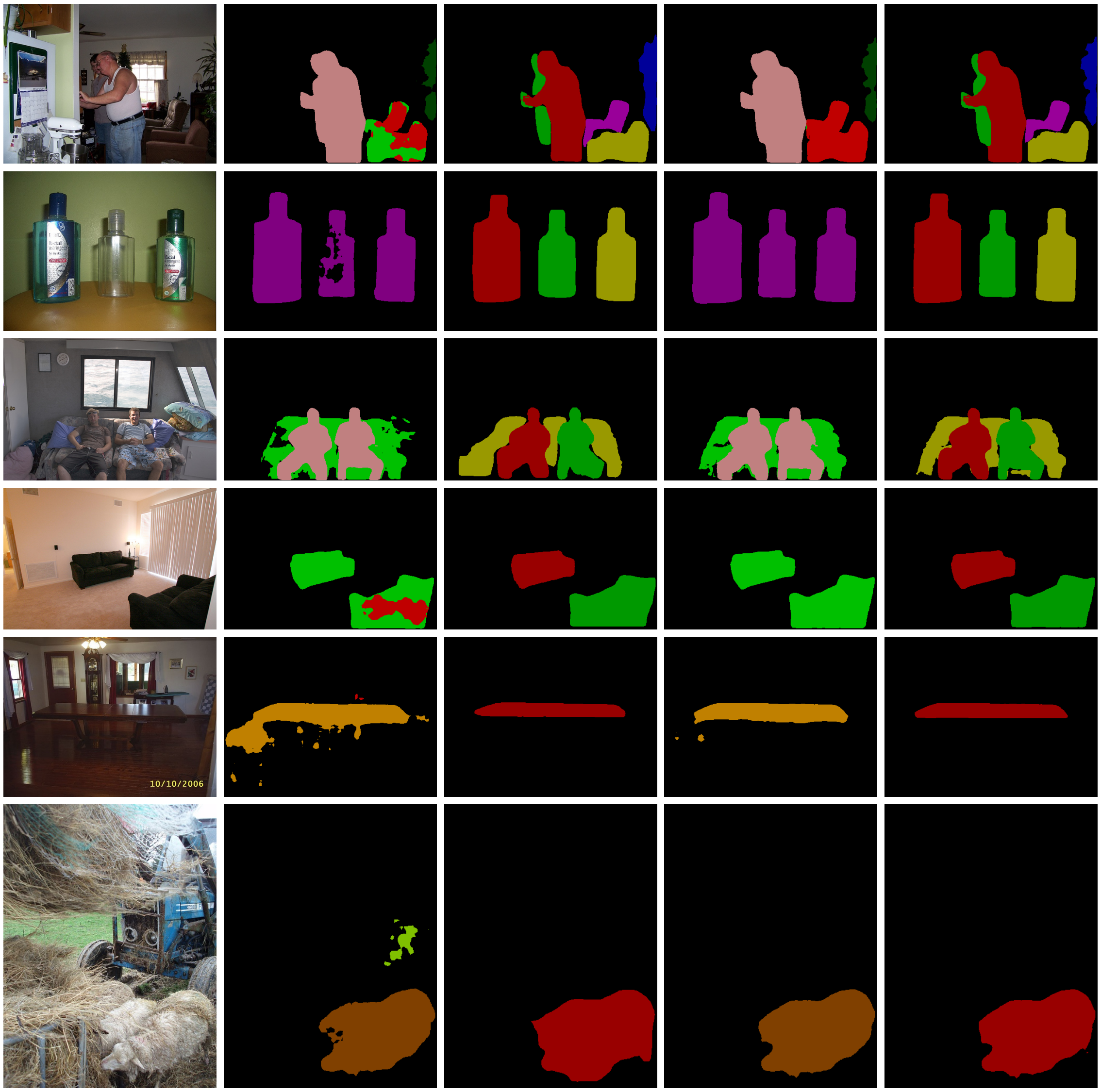

Table 3.5: **Visualizations on Pascal VOC.** Example images from the Pascal VOC validation set. Columns left to right: original image, semantic output before BCRF, instance output before BCRF, semantic output after BCRF, instance output after BCRF. Each row contains a new image. The standard Pascal VOC color map is used for the semantic segmentation results.

	PQ		**SQ**		**RQ**		
Category	W/O BCRF	BCRF	W/O BCRF	BCRF	W/O BCRF	BCRF	Classes
All	41.4	41.7	78.3	79.1	50.8	51.1	133
Things	47.4	47.4	80.4	80.4	57.3	57.3	80
Stuff	32.5	33.2	75.1	77.1	40.9	41.6	53

Table 3.6: **COCO dataset.** Panoptic segmentation results on the COCO validation set.

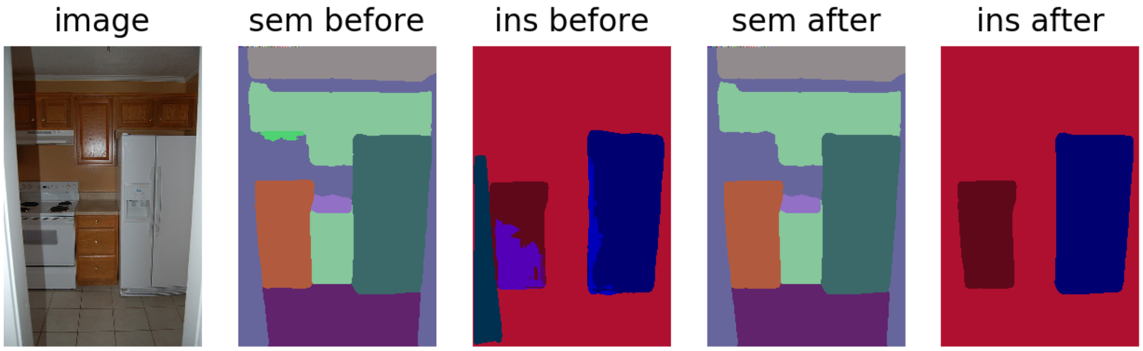

Figure 3.2: Visualisation of improvements on COCO Dataset

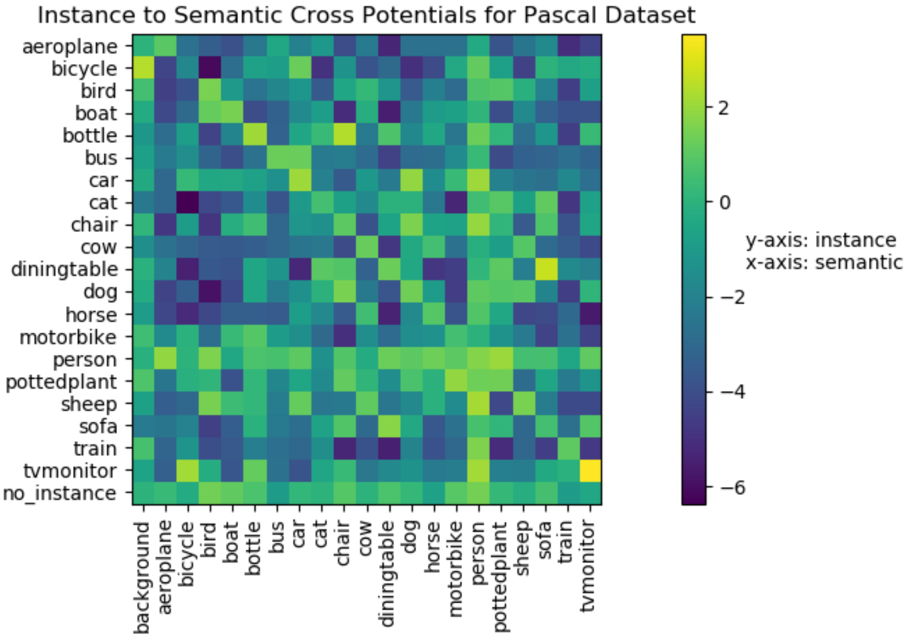

Figure 3.3: The heatmap illustrates inter-class dependencies learned by the cross-potential term weights of BCRF. Note that a logarithmic scale has been used.

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 [3]. 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 [1]. However in [1] 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 [19] 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 [15] for training.

References

[1] 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

[2] A. Geiger, P. Lenz, and R. Urtasun, "Are we ready for autonomous driving? the kitti vision benchmark suite," in *Conference on Computer Vision and Pattern Recognition (CVPR)*, 2012.

[3] 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.

[4] A. Bewley, Z. Ge, L. Ott, F. Ramos, and B. Upcroft, "Simple online and realtime tracking," in *IEEE International Conference on Image Processing (ICIP)*, 2016, p. 34643468.

[5] A. Milan, S. H. Rezatofighi, A. Dick, I. Reid, and K. Schindler, "Online multi-target tracking using recurrent neural networks," in *Conference on Artificial Intelligence, Association for the Advancement of Artificial Intelligence (AAAI)*, 2016.

[6] Y. Xiang, A. Alahi, and S. Savarese, "Learning to track: Online multi-object tracking by decision making," in *International Conference on Computer Vision (ICCV)*, 2015, p. 47054713.

[7] J. Zhu, H. Yang, N. Liu, M. Kim, W. Zhang, and M.-H. Yang, "Online multi-object tracking with dual matching attention networks," *Computer Vision ECCV*, vol. 11209, p. 379396, 2018.

[8] J. Kuck, "Target tracking with kalman filtering," in *ArXiv*, 2016.

[9] M. Kim, S. Alletto, and L. Rigazio, "Similarity mapping with enhanced siamese network for multi-object tracking," in *Machine Learning for Intelligent Transportation Systems (MLITS)*, 2016.

[10] P. H. T. Vibhav Vineet, Jonathan Warrell, "Filter-based Mean-Field Inference for Random Fields with Higher-Order Terms and Product Label-Spaces," in *ECCV*, 2012.

[11] Koller, Daphne and Friedman, Nir, *Probabilistic Graphical Models: Principles and Techniques - Adaptive Computation and Machine Learning*. The MIT Press, 2009.

[12] P. Krähenbühl and V. Koltun, "Efficient Inference in Fully Connected CRFs with Gaussian Edge Potentials - Supplementary Material," in *NIPS*, 2011.

[13] Y. Xiong, R. Liao, H. Zhao, R. Hu, M. Bai, E. Yumer, and R. Urtasun, "Upsnet: A unified panoptic segmentation network," in *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2019.

[14] J. Long, E. Shelhamer, and T. Darrell, "Fully Convolutional Networks for Semantic Segmentation," in *CVPR*, 2015.

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

[16] T.-J. Yang, M. D. Collins, Y. Zhu, J.-J. Hwang, T. Liu, X. Zhang, V. Sze, G. Papandreou, and L.-C. Chen, "Deeperlab: Single-shot image parser," *ArXiv*, vol. abs/1902.05093, 2019.

[17] A. Kirillov, K. He, R. B. Girshick, C. Rother, and P. Dollár, "Panoptic segmentation," in *CVPR*, 2019.

[18] J. Huang, V. Rathod, C. Sun, M. Zhu, A. K. Balan, A. Fathi, I. Fischer, Z. Wojna, Y. Song, S. Guadarrama, and K. Murphy, "Speed/accuracy trade-offs for modern convolutional object detectors," *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 3296–3297, 2016.

[19] 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.

[20] 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.

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

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 [20], where $\theta_{pred}, \theta_{gt}$ are predicted and ground truth values of the parameter θ respectively.

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

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