Recognition and Segmentation of Surgical Gestures

Supervisor : PD Dr.habil.Rudolph Triebel
Mentor : Antonin Raffin
Author : Pooja Krishnan

DLR

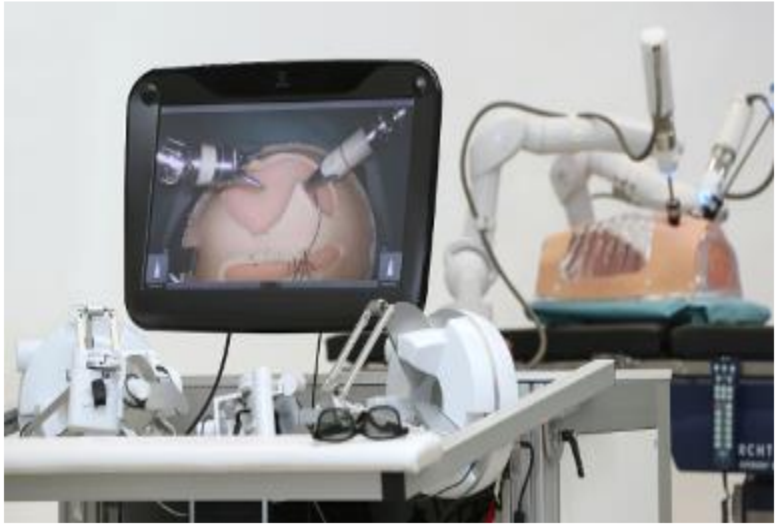
Outline

- Motivation
- Problem statement
- Datasets – JIGSAWS and MIROSurge
- Model – Bidirectional LSTM
- Uncertainty Estimation
- Methods to quantify uncertainty
- Results
- References

Motivation

Why minimally invasive surgery?

- Less pain and blood loss
- Faster recovery and smaller, less noticeable scars
- Lesser chances of surgical site infection
- Reduction in surgery time
- Improvement in quality of surgical procedures.



Problem Statement

- Recognition and segmentation of surgical gestures using video data, kinematic data and both.
- Confidence estimation for each gesture.
- Top-3 most probable gestures.
- Evaluated on JIGSAWS and MIROSurge Dataset.

Datasets

JIGSAWS

Surgical Tasks:

- Suturing
- Knot Tying
- Needle Passing

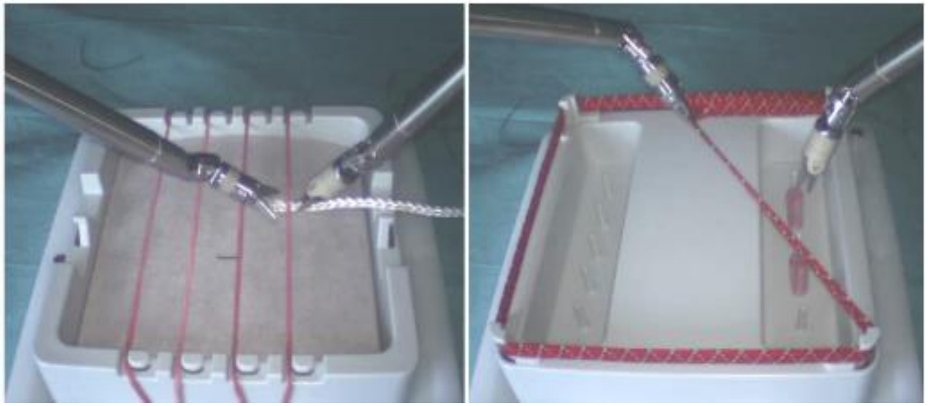
- 8 users, 5 trials each
- Recorded at 30Hz
- 76 kinematic features

MIROSurge

Surgical Tasks:

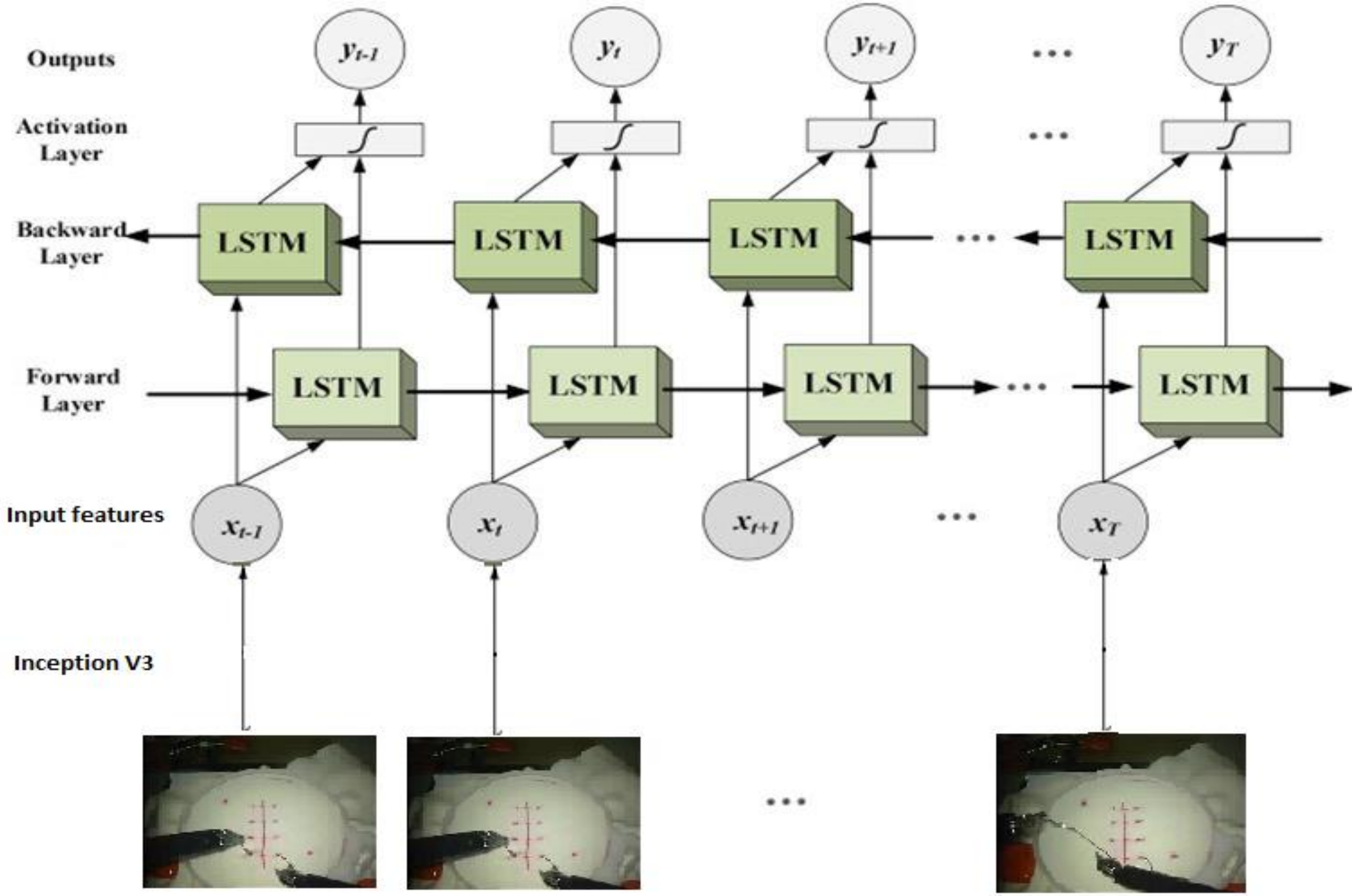
- Band Twist
- Weaving

- 4-5 users, 3 trials each
- Recorded at 30Hz
- 49 kinematic features



DLR

Model – Bidirectional LSTM



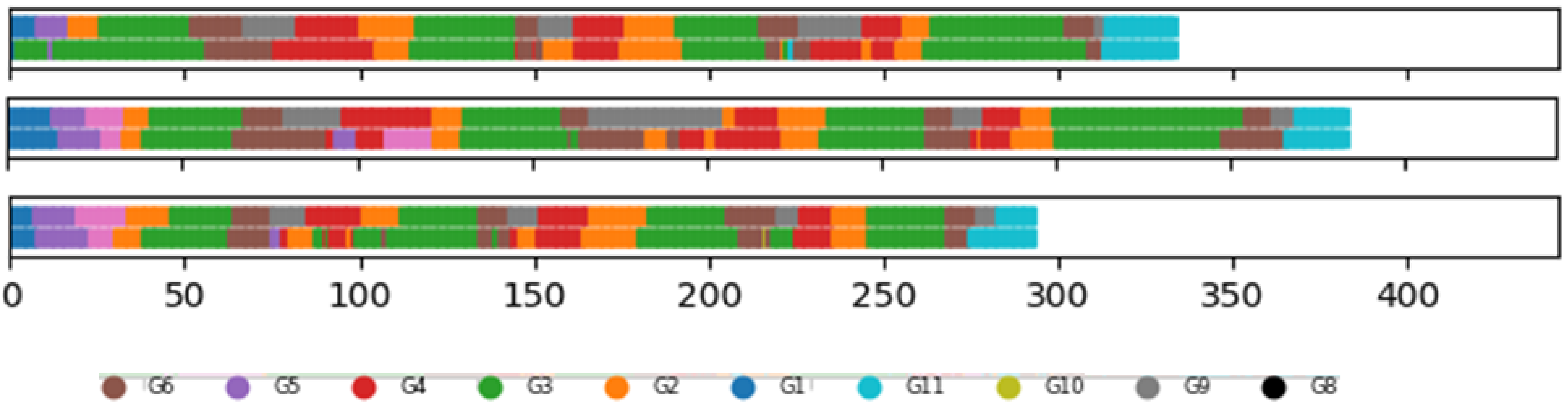
Source: adapted from 'A novel wavelet sequence based on deep bidirectional LSTM network model for ECG signal classification'

Results – JIGSAWS – Video Data

Timeline Visualization- Suturing Task on Video data (LOUO)

Data Pre-processing:

- Add variable gaussian noise every epoch
- Augmentation: Crop, Blur ; Scaling, Translation



- The overall accuracy is affected my unseen gestures. For example, User 'D' performs gesture ' 'G9' often unlike other Users
- Oversegmentation and offsets
- Metrics: Accuracy and Edit Distance

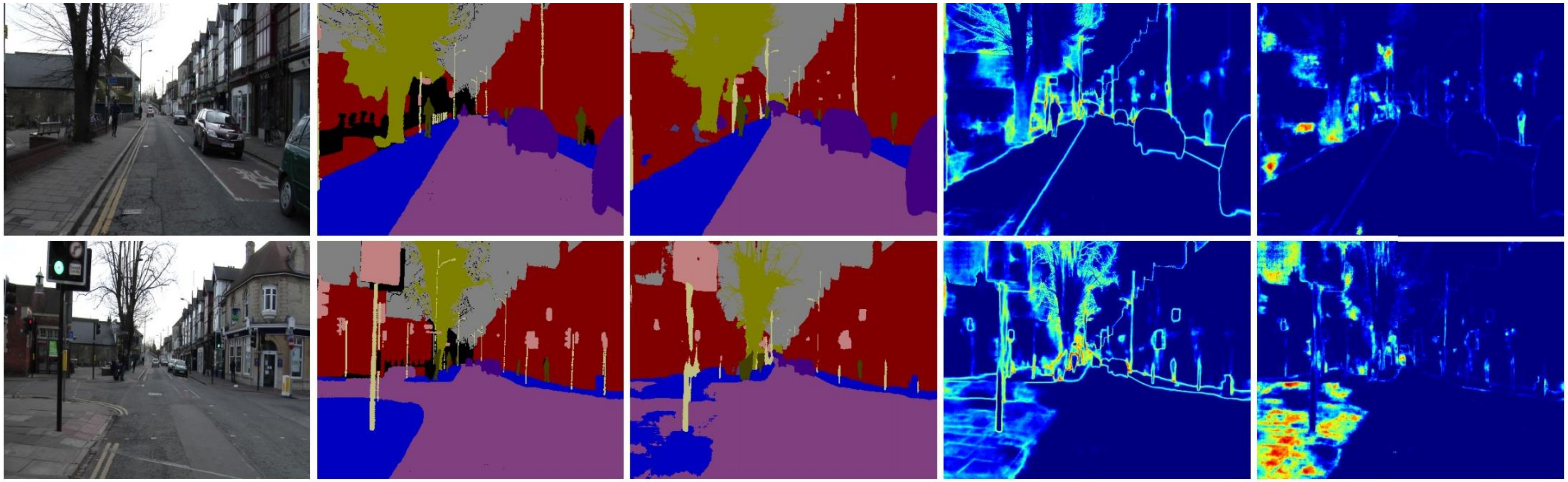
Uncertainty Estimation

Epistemic uncertainty (*model uncertainty)*

- Accounts for uncertainty in the model parameters
- Important to model for Safety-critical applications and small datasets where the training data is sparse.

Aleatoric uncertainty*(data or task dependent)*

- Captures noise inherent in the observations
- Important to model for large data situations and real time applications

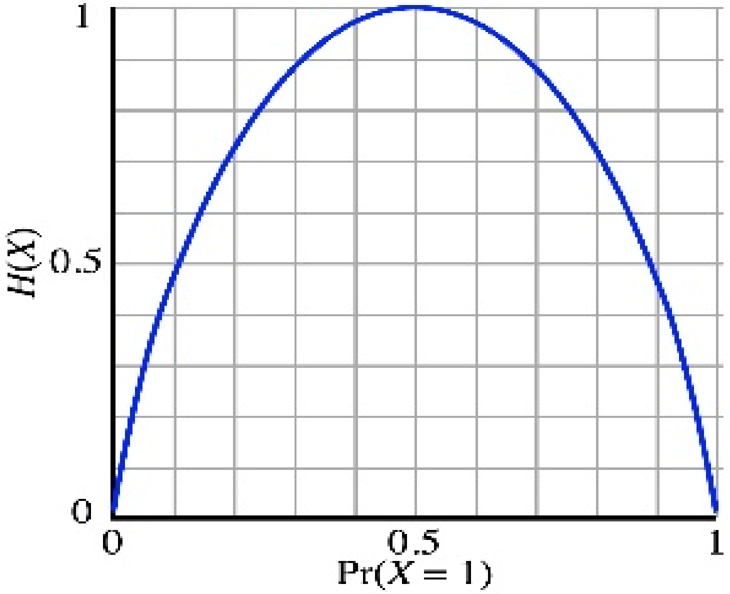


(a) Input Image (b) Ground Truth (c) Semantic Segmentation (d) Aleatoric Uncertainty (e) Epistemic Uncertainty

Uncertainty Estimation- Classification Task

Predictive Entropy

$$H[\, y \mid x, D_{train}] = -\sum_c p(y = c | x, D_{train}) \log p(y = c | x, D_{train})$$



- Predictive Entropy is maximum, when all classes have equal probability
- Predictive Entropy is 0, when one class has probability of 1.

Methods to quantify uncertainty

Monte Carlo Dropout

Bernoulli approximate variational inference in Bayesian NN can be achieved by adding dropout during training and test time.

Drawback:

Test time is scaled by a factor of the number of averaged forward passes through the network.


GT:G1—Reaching for needle with right hand
PD:G1—Reaching for needle with right hand(46.14%)

Results – JIGSAWS – Uncertainty Estimation

Naive Ensembles

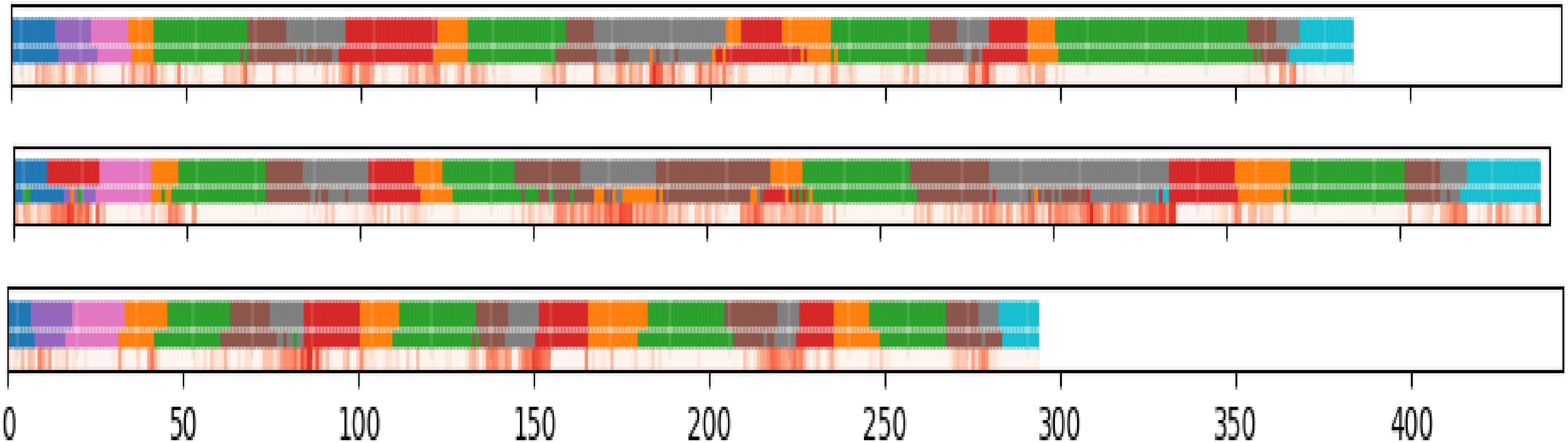
Algorithm 1: Uncertainty Estimation using Naïve Ensemble

Result: Uncertainty estimation

Initialize the parameters $\theta_1, \theta_2, \theta_m$ with random values;

Train each of the $m \epsilon M$ networks with the θ_m values;

Combine the predictions: $p[y|x] := \frac{\sum_{m=1}^{M} p_{\theta_m}(y|x, \theta_m)}{M}$



Result for classification tasks

Ensemble with Bootstrap

Algorithm 2: Uncertainty Estimation using Bootstrap Ensemble

Result: Uncertainty estimation

Select one user of M users as the test user;

Train each of the $m \epsilon M$ models by leaving one user out;

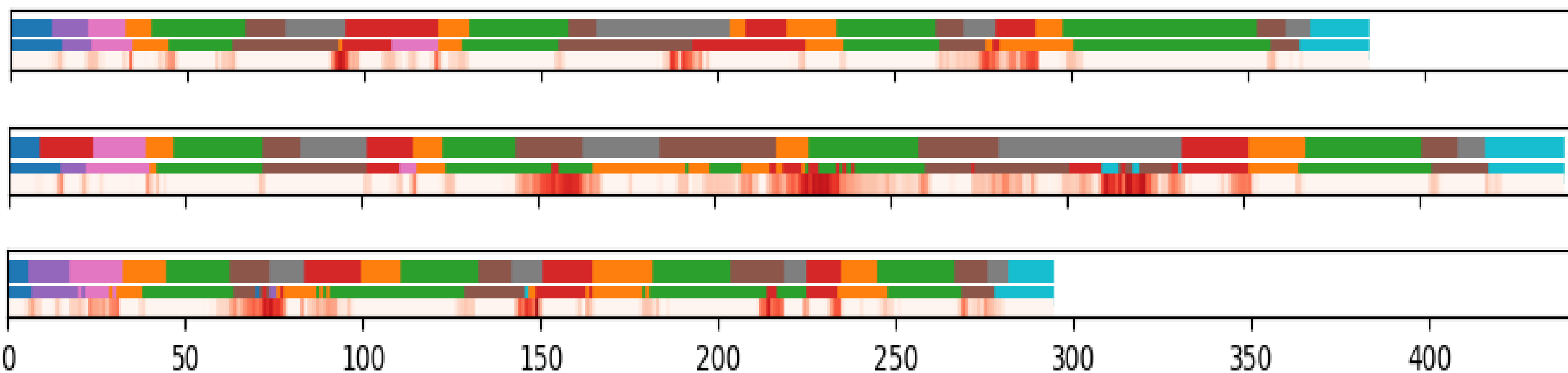
for *m in 1,..M-1* **do**

$\quad P(y|x) := f_{\theta m}(x)$

end

$$P(y=c|x) = \frac{\sum_{m=1}^{M-1} \sum_c P(y|x)}{(M-1)};$$

$$H[y|x, D_{train}] := -\sum_c p(y=c|x) log p(y=c|x);$$



Results – JIGSAWS – Video Data

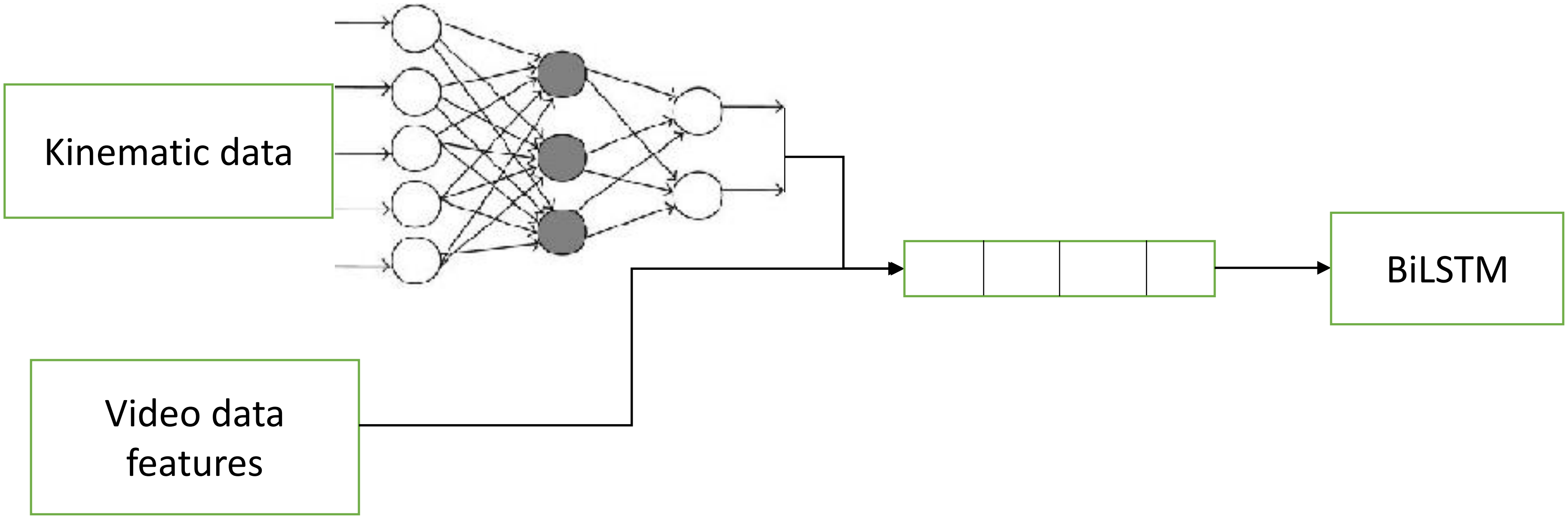
Comparison with state-of-the-art						
Models	Suturing		Knot Tying		Needle Passing	
	Accuracy	ED	Accuracy	ED	Accuracy	ED
CRF[19]	68.8	NA	60.17	NA	54.52	NA
Semi CRF[19]	59.41	NA	41.46	NA	46.89	NA
MsM CRF[19]	71.76	NA	66.94	NA	60.39	NA
Bidirectional LSTM	76.2	12.14	70.44	23.33	49.95	85.18
ED-TCN[15]	81.4	11.1	NA	NA	NA	NA
TDRN[16]	84.6	10.2	NA	NA	NA	NA

Results for LOUO

Comparison with state-of-the-art						
Models	Suturing		Knot Tying		Needle Passing	
	Accuracy	ED	Accuracy	ED	Accuracy	ED
Semi CRF[19]	65.83	NA	44.82	NA	56.22	NA
CRF[19]	76.51	NA	69.16	NA	62.23	NA
MsM CRF[19]	79.04	NA	72.04	NA	68.81	NA
Bidirectional LSTM	81.17	8.078	76.24	19.014	56.73	50.36

Results for LOTO

Results – JIGSAWS – Kinematic+ Video Data



Comparison with state-of-the-art						
Models	Suturing		Knot Tying		Needle Passing	
	Accuracy	ED	Accuracy	ED	Accuracy	ED
MsM-CRF(kin-STIP)[19]	70.09	NA	68.43	NA	54.41	NA
MsM-CRF(kin-dense)[19]	72.6	NA	68.83	NA	57.08	NA
MLP+Bidirectional LSTM	81.38	10.51	75.57	20.33	68.55	49.42

Results – JIGSAWS – Kinematic Data

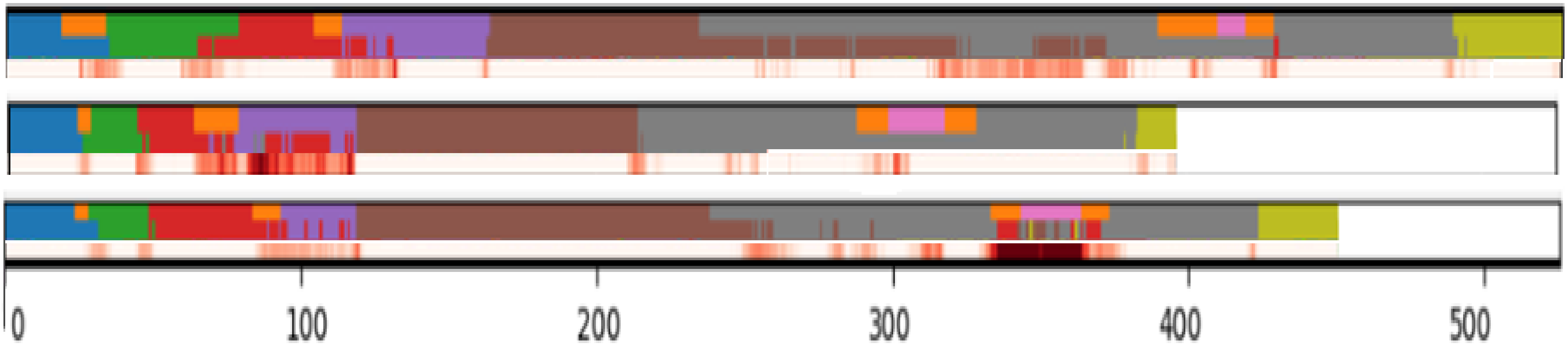
Comparison with state-of-the-art						
	Suturing		Knot Tying		Needle Passing	
Models	Accuracy	ED	Accuracy	ED	Accuracy	ED
MsM-CRF[19]	69.03	NA	64.28	NA	52.39	NA
GMM-HMM[19]	73.95	NA	64.13	NA	72.47	NA
Forward LSTM	80.5	NA	NA	NA	NA	NA
Bidirectional LSTM	82.03	6.24	83.15	9.5	77.21	15.82
MS-RNN	90.2	NA	NA	NA	NA	NA

Results – JIGSAWS – Top-3 Accuracy

Top-n accuracy						
	Top-1		Top-2		Top-3	
Tasks	LOUO	LOTO	LOUO	LOTO	LOUO	LOTO
Suturing -vid	76.2	81.176	89.96	92.30	93.435	94.789
Suturing -vid+kin	81.38	86	91.537	93.73	93.61	95.23
Suturing -kin	82.03	87.09	91.86	94.67	94.39	96.38
Knot Tying -vid	70.44	76.24	75.63	89.88	92.16	93.79
Knot Tying -vid+kin	75.57	82.052	89.84	92.9	94.93	96.36
Knot Tying -kin	83.15	86.5	93.57	94.84	96.67	96.84
Needle Passing -vid	49.95	65.12	70.458	80.51	81.18	90.2
Needle Passing -vid+kin	68.5	75.95	84.74	88.3	90.18	91.67
Needle Passing -kin	77.21	79.61	88.96	89.76	92.29	92.44

Results – MIROSurge Dataset

Comparison of Methods				
Methods	Band Twist		Weaving	
	LOUO	LOTO	LOUO	LOTO
MPL C[12] -kin	17	25	31	42
Bidirectional LSTM - kin	35.15	39.23	62.51	65.2
Bidirectional LSTM - kin+vid	68.052	76.51	73.95	75.12
Bidirectional LSTM - vid	66.52	72.54	71.38	73.2



Timeline visualization based on kinematic data

References

- G. et al. The JHU-ISI gesture and skill assessment working set (JIGSAWS): A surgical activity dataset for human motion modeling. Modeling and Monitoring of Computer Assisted Interventions (M2CAI) – MICCAI Workshop, 3, 2014.
- https://www.dlr.de/rm/en/desktopdefault.aspx/tabid-11674/#gallery/28728
- https://alexgkendall.com/computer_vision/bayesian_deep_learning_for_safe_ai/

Extra Slides!!!