

Dexterous Skill Transfer between Surgical Procedures for Teleoperated Robotic Surgery

Mridul Agarwal¹, Glebys Gonzalez², Mythra V. Balakuntala³, Md Masudur Rahman⁴,
Vaneet Aggarwal^{1,2}, Richard M. Voyles³, Yexiang Xue⁴, and Juan Wachs²

Abstract—In austere environments, teleoperated surgical robots could save the lives of critically injured patients if they can perform complex surgical maneuvers under limited communication bandwidth. The bandwidth requirement is reduced by transferring atomic surgical actions (referred to as “surges”) instead of the low-level kinematic information. While such a policy reduces the bandwidth requirement, it requires accurate recognition of the surges. In this paper, we demonstrate that transfer learning across surgical tasks can boost the performance of surge recognition. This is demonstrated by using a network pre-trained with peg-transfer data from Yumi robot to learn classification on debridement on data from Taurus robot. Using a pre-trained network improves the classification accuracy achieves a classification accuracy of 76% with only 8 sequences in target domain, which is 22.5% better than no-transfer scenario. Additionally, ablations on transfer learning indicate that transfer learning requires 40% less data compared to no-transfer to achieve same classification accuracy. Further, the convergence rate of the transfer learning setup is significantly higher than the no-transfer setup trained only on the target domain.

I. INTRODUCTION

Surgical debridement is a key surgical skill necessary when burn injuries occur on the battlefield. The prime surgical task is to remove dead tissues from the skin and allow healthy tissue to heal (Figure 1). Timely interventions to patients can help provide initial treatment before the patients can be evacuated to more and better equipped surgical centers. To achieve this timely intervention, there is an increasing interest in using teleoperated surgical robots [1], [2] to support such tasks. Yet, so far the performance of teleoperated robotic systems is susceptible to bandwidth and latency of the underlying communication network [3]. Thus, there is a need for platforms with semi-autonomous capabilities that can assist the surgeon (or the medic) when communication is hindered.

Due to the vast amount of information involved in teleoperation, it is more efficient to use a high-level representation of the environment, patient and robot, to accomplish this task. These main pieces of information would be enough to predict the action being performed by the teleoperating surgeon and complete it autonomously by the robot. In

order to effectively interpret the environment, it is necessary to recognize the current and previous surgical actions, referred to as ‘surges’ [4]. Nevertheless, surgical data for austere environment procedures is scarce [5], hindering the training networks that can automatically recognize these surgical actions. Moreover, field medical robots are diverse, holding different kinematic configurations, workspaces, and operate under partially unknown constraints. Such domain differences could hamper state-of-the-art approaches and prevent models learned on one platform to generalize across other platforms of disparate morphologies [9]. To tackle these domain limitations, we propose a framework that can leverage the existing abundant surgical laparoscopic datasets and transfer it across different robots and tasks. Thus we propose a framework for surge recognition that can be trained in one domain (peg-transfer using the Yumi robot) and successfully recognize surges in another domain (debridement using the Taurus robot).

The framework aims to leverage the abundance of data available from more accessible environments to find recurrent patterns and apply such insights to new scenarios [6]–[8]. To transfer knowledge between surgical tasks, the surge classes of the laparoscopic peg-transfer are mapped to the surge classes of the surgical debridement. Then an LSTM-based architecture is trained using sequences of kinematic features from a peg-transfer task. Once the classifier learns to recognize surges in the source domain (peg-transfer task), it is not required to be trained from scratch in the target domain (debridement task) but initialized with source domain parameters. Reusing the knowledge learned from the source domain features significantly reduces the amount of training required from the target domain.

To assess the performance of the framework, the surge classification accuracy was obtained in two different setups. The first one is a no-transfer baseline, where the model is trained solely on the target domain (the debridement dataset). This setup produced a frame-wise classification accuracy of 83%. Then, the recognition was tested on the transfer learning setup, producing a classification accuracy of 85%. Lastly, an ablation study was conducted, showing that the transfer-learning setup’s convergence rates are faster than the baseline (no-transfer setup).

The contribution of this paper can be summarized as follows:

- 1) A dataset for training classifiers for the debridement

¹School of Electrical and Computer Engineering, Purdue University, West Lafayette, IN, 47907, USA agarw180@purdue.edu

²School of Industrial Engineering, Purdue University, West Lafayette, IN, 47907, USA gonza337, vaneet, jpwachs@purdue.edu

³School of Engineering Technology, Purdue University, West Lafayette, IN, 47907, USA mbalakun, rvoyles@purdue.edu

⁴Department of Computer Science, Purdue University, West Lafayette, IN, 47907, USA rahman64, yexiang@purdue.edu



Fig. 1: Human performing debridement task on a burnt skin model. The task involves manipulating the dead skin, cutting it, and discarding the dead skin. The healthy tissue is exposed after the cut surgeme.

task using the Taurus robot¹.

- 2) A transfer learning framework to learn surgemes for a debridement task using a small sized dataset. We show that using a transfer learning setup allows to reduce data requirement by 32% to obtain similar performance on test setup. Further, for smaller datasets, transfer learning works with upto 15% higher classification accuracy.
- 3) An ablation study which helps in understanding which layers contribute the most to the learning tasks. We show that simply re-training the softmax layer is not sufficient, indicating that retraining the layers which capture the temporal abstractions of the input data is required.

The rest of this paper is presented as follows: Section II discusses the prior work. Section III gives an overview of the robotic dataset used. Section IV describes the methods used for surgeme classification using LSTM. Section V describes the transfer learning setup using peg transfer. Section VI shows the experimental setup, ablation studies, results, and discussion. Finally, Section VII concludes the paper with a discussion on future work.

II. BACKGROUND AND RELATED WORK

Datasets play a crucial role in developing automated detection tools for surgical robotic activity [6], [7]. The advantage of these datasets is that they allow elucidating patterns associated with skill learning where surgical tasks are decomposed into a finite set of maneuvers [6]. In this decomposition, each surgical skill is represented as a sequence of atomic units referred to as surgemes. This process of decomposition is known as surgical skill modeling [9].

In the past, various methods have been proposed for surgeme recognition including using Hidden Markov Models [10]–[13], structured prediction [14], [15], and recurrent neural networks [8]. All of these use the JIGSAWS dataset [6] which consists of three surgical procedures: suturing, knot tying, and needle passing, performed with the da Vinci Surgical System. However, JIGSAWS is recorded on a single robotic platform, whereas the DESK dataset [7] has data

from different robotic platforms. In DESK, the peg-transfer task was performed on various robotic platforms, including simulators (Taurus, da Vinci, Yumi). This dataset serves as a testbed for surgeme prediction, including transfer learning in surgical activity recognition [7], [16]–[19]. These works focus on transfer learning across different robotic platforms, but the data is from the same domain (peg transfer).

Transfer learning is a method in machine learning, where the training data and the testing data may not be identically distributed [20]. For transfer learning, the classifier for the target domain is not required to be trained from scratch but initialized with the parameters trained for another domain (known as source domain). A known method of achieving transfer learning between domains, is to train a network in the source domain and then retrain the last layers of this network in the target domain [21], [22]. This procedure often allows boosting the classification accuracy in the target domain with less data, since the network can benefit from the learnt low-level statistics and does not have to start training from scratch [23], [24]. This methodology has been successfully used in computer vision and natural language processing tasks [25]–[27]

In the field of surgical robotics, similar method for re-training has been used to leverage the information from known datasets. Zhang et al. [28] uses the JIGSAWS dataset to pre-train a network for skill assessment. The authors present a model of 1D-CNNs that is trained using samples from JIGSAWS. This process accumulates feature information in the first two layers, and the last few layers are re-trained on the Robot-Assisted Microsurgery (RAMS) dataset. The work of Tsai et al [29] proposes another transfer learning architecture for surgeme segmentation that leverages the JIGSAWS dataset. Their method uses a Self Similarity Matrix (SSM) as a feature extractor, that is in turn fed as the state input to train a policy on the source domain.

As surgemes are examples of sequential data, re-training only the last few layers might be ineffective. Thus, we analyze the effect of retraining the entire network or only few layers.

¹<https://bit.ly/3fJzklC>

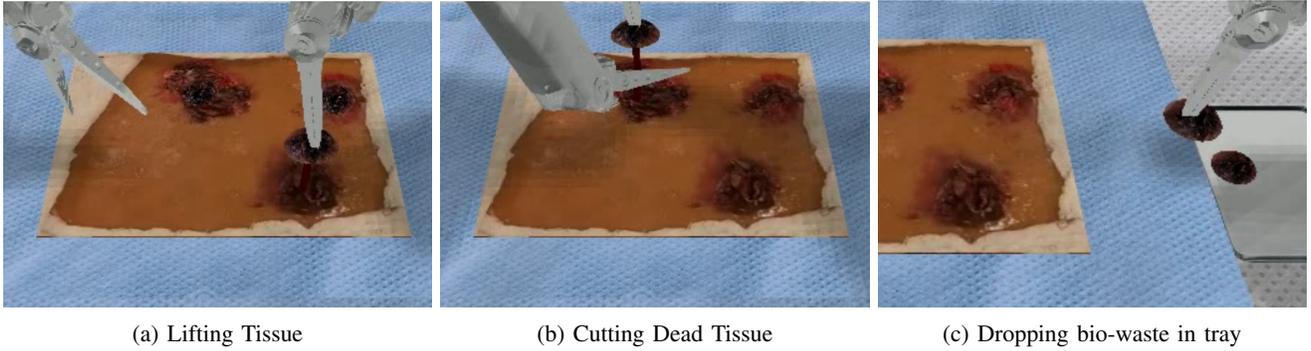


Fig. 2: The debriement task performed inside the CopelliaSim simulator using a Taurus robot. The tool tips of the Taurus robot is modified to lift or cut the dead skin by acting like dissection scissors.

III. THE DEBRIDEMENT TASK

The debriement task requires removal of dead tissue or dead tissue fragments to allow the underneath healthy tissue to heal [30]. This requires manipulating the robotic arms such that it grasps the diseased tissue, incise it, and then safely discard the bio-waste. In light of these task, we divide a debriement task robotic maneuver into 7 surgemes (labelled as $S_i, i \in \{1, \dots, 7\}$) as follows:

- **S1: Approach Skin** requires to move the robot arm towards the dead skin.
- **S2: Align and Grasp** requires to hold the dead skin in the most appropriate way.
- **S3: Lift** requires to lift the dead tissue to allow to cut it.
- **S4: Approach Blade** requires to bring the surgical blade close to the lifted tissue.
- **S5: Cut** requires to ensure slice the dead tissue to detach it from the rest of the body. Depending upon the requirements, there might be multiple slices required to properly cut the dead tissue.
- **S6: Approach Bin** requires to move the arm holding the removed dead tissue towards the tray.
- **S7: Drop tissue** requires to safely and completely drop the dead tissue in the tray.

Figure 1 shows the dead removed and the healthy tissue exposed by the debriement task on artificial skin.

A. Dataset

We note that it is extremely difficult to obtain or produce a real-world dataset of a debriement task as the one would require to create skin modes as shows in Figure 1. Hence, we developed a simulator to perform a debriement task using CopelliaSim.

To model the elasticity of skin in the debriement task, we use springs within the simulated skin. In the simulation we have 3 patches of dead skin which must be removed to successfully complete the surgical task. Further, after removing the necrosed tissue, it must be deposited over a surgical tray. We use a Taurus robot as our execution robot inside the simulator. We used the Taurus robot with tool tips in the form dissection scissors to cut through the dead



Fig. 3: Interface for robot control

skin. The robot is controlled using Razer Hydra sensing controller which allows a natural movement for operator's hand throughout the task, as shown in figure 3.

The collected kinematics includes the robot end-effector position in Cartesian coordinates from the robot's frame of reference and orientation angles for both, the left and the right arms. Additionally, it also includes the state of the gripper to include whether the gripper is in open or in closed position. This schema results a kinematic sequence where the kinematic vectors consists of $3 + 4 + 1$ elements for each arm resulting in 16 elements for both arms.

We collected and annotated data from 3 subjects. Each subject completed 12 trials of a debriement task and each debriement task requires removing all 3 dead tissue patches from the skin. Each subject conducted 6 trials. In each trial, it is required to grasp the necrosed tissue from the left robot arm and cut the skin from the right robot arm. Similarly, they also completed 6 trials where the roles of the arms were inverted to allow the model to learn ambidexterity.

IV. CLASSIFICATION OF SURGEMES FOR DEBRIDEMENT TASK

The operator interacting with the simulator generates a sequence of kinematics corresponding to a surgeme. To classify the kinematic sequence, we use an LSTM based neural network [31] was used in combination with the neural

network proposed in [32].

We use the simulated robot kinematics as the input to the classifier. Let $\{\mathbf{x}_t\}_{1 \leq t \leq T}$ be the input kinematic sequence where \mathbf{x}_t is the simulated robot kinematics at time t , and T is the length of the surgical task. Note that a human operator generates the sequence data in an online manner instead of generating complete sequence. This requires to make the surgeme classification task to be causal as it cannot depend on future kinematics to predict the current surgeme. To make the neural network classifier causal, we pass the kinematics \mathbf{x}_t as the input to the LSTM at time t . Further, as the length of each each surgeme is variable, we cannot use methods for fixed length sequences such as 3D-CNN [7], [16], [33].

In our network architecture, the LSTM layer is followed by a dense network to compute logit values. We used softmax activation on logits to classify the input sequence into the 7 surgemes. Let the true labels be $\{\mathbf{y}_t\}_{t=1}^T$ where $\mathbf{y}_t \in \Delta^K$ and Δ^K is a probability simplex in K dimensions for K classes (for 7 surgemes $K = 7$). The LSTM is used to predict the class probability $\hat{\mathbf{y}}_t \in \Delta^K$ at time t . We use a cross entropy loss defined in Equation (1) to train the LSTM network.

$$L(\mathbf{y}, \hat{\mathbf{y}}) = \sum_{t=1}^T \sum_{k=1}^7 \mathbf{y}_{t,i} \log(\hat{\mathbf{y}}_{t,i}) \quad (1)$$

To implement the described network, we configured the LSTM network with a hidden state size of 32 followed by a fully connected layer of size 32×7 . The final softmax layer returns the probabilities distribution of the current surgeme given the current kinematic input. The surgeme that will be eventually executed is the one with the highest probability given the input kinematic \mathbf{x}_t .

V. TRANSFER LEARNING FOR SURGEME RECOGNITION

We used a pre-trained network for classifying the surgemes in a peg transfer task to use as initial weights for the network described in previous section. We use the classifier model of [32] use for surgeme identification in the peg and pole task. The network structures for the two tasks are explained in Figure 5.

The peg transfer task and the debridement task involve similar surgemes such as approach skin or approach peg and align and grasp, however, the kinematic sequences may be drastically different. For example, the peg-and-pole task requires the peg manipulation using both robotic arms one after the other. However, in the debridement task, one arm only acts as the blade and the other arms holds the object for the entire duration of debridement. To demonstrate that the two sequences are different, we use dynamic time warping (DTW) distance [34]. Table I shows the average DTW distance between any two sequence for the two surgical tasks.

	Peg Transfer	Debridement
Peg Transfer	96.25	112.69
Debridement	112.69	106.24

TABLE I: Average Dynamic Time Warp distances between the sequences of the two surgical task.

Using pre-trained networks allows for not only training with less amount of data, but it also allows for an instantaneous use of the pre-trained network with few runs (< 50) of back-propagation steps. We also note that the temporal information is extracted by the LSTM layer, and the two sequences for the debridement task and the peg and pole task are different sequences, we must retrain the LSTM layer. In the next section we present the effect of re-training and freezing the layers of the surgeme classification network, impact of increasing the training data of the debridement task, and the rate to convergence of the transfer learning.

VI. RESULTS AND DISCUSSION

First the results without transfer learning is presented, followed by the results from the different transfer learning setups. The proposed framework in Fig. IV is evaluated on a dataset of 88 surgeme sequences from 3 subjects with 12 trials each. The dataset is separated with a test-train split of 70% and 30% respectively. The surgeme recognition results are presented in Table II.

Accuracy	S1	S2	S3	S4	S5	S6	S7	Average
Testing	97	48	76	84	83	85	87	83.7
Training	97	59	83	93	88	88	89	90.4

TABLE II: Surgeme classification accuracy of the baseline with non-transfer learning.

As the proposed network few parameters to learn, 88 sequences are sufficient to achieve good classification. However, transfer learning achieves a similar performance with even lesser training sequences as shown in the following section.

A. Transfer learning results

First, a direct transfer learning approach is adopted. The network is trained with data from peg-transfer task. The peg-transfer dataset was collected using 5 subjects with 12 trails per subject resulting in a total of 185 sequences. After a 70% – 30% train-test split and training the network with the resulting 130 sequences, an average accuracy of 77% was obtained. This network was re-trained on the debridement data with the same split as the no-transfer scenario. The resulting surgeme classification accuracy is shown in Table III.

Accuracy	S1	S2	S3	S4	S5	S6	S7	Mean
Baseline	97	48	76	84	83	85	87	83.7
Transfer Learning	94	63	74	86	84	79	91	85.8

TABLE III: Surgeme classification accuracy on the test dataset for the baseline (no-transfer learning) and transfer learning methods.

When the entire dataset is used to train the proposed model, the transfer learning setup performed marginally better than the no-transfer setup. However, the transfer learning setup performed significantly better than the no-transfer setup as shown in the following section.

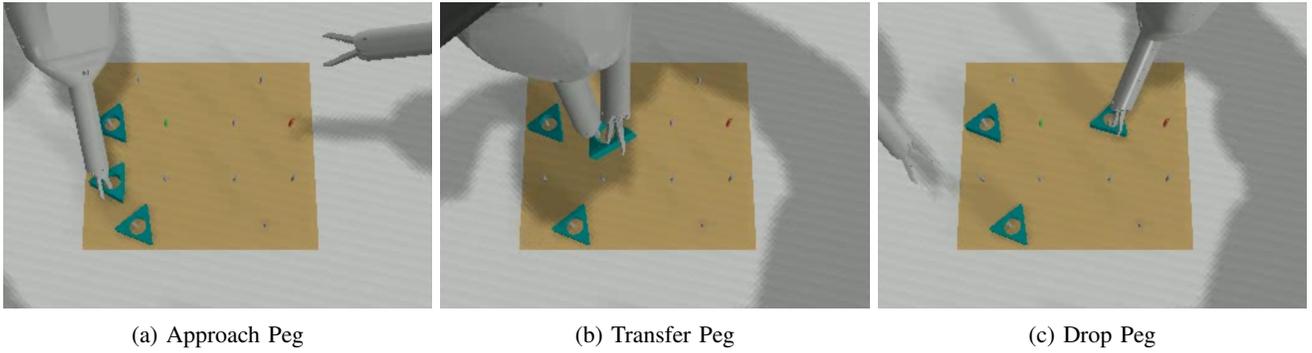
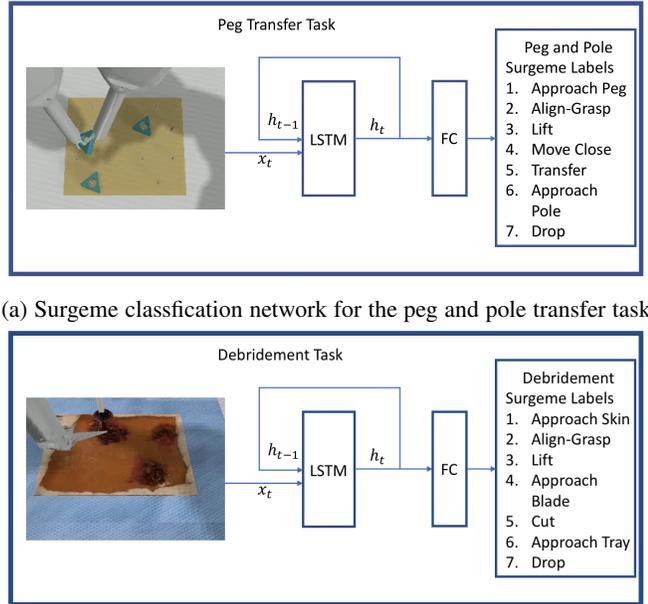


Fig. 4: The peg transfer task used for transferring skills to the surgeme classification for the debridement task.



(a) Surgeme classification network for the peg and pole transfer task

(b) Surgeme classification network for the debridement task

Fig. 5: The fundamental structure of the networks is kept same to allows for transfer of the knowledge between the two classifiers. First the network is trained on the Peg and Pole task where both the arms are used extensively. Pre-trained network is then used to transfer to the new debridement task. Since the number of classes are the same, we do not modify the structure of the last layer.

B. Transfer Learning Analysis

The get more insight into the transfer learning setup, three different aspects of the method are analyzed in the following sub-sections,

1) *Impact of Amount of Training Data:* The debridement dataset has 88 sequences of the task out of which 61 sequences were separated as training set. The remaining 17 sequences are used as the test set. To study the impact of the amount of training data a fraction of the 61 sequences is used for training and tested with the same 17 sequences. Different fractions starting from 10% to 70% (61 sequences) in steps of 10% were used. For the transfer learning, the

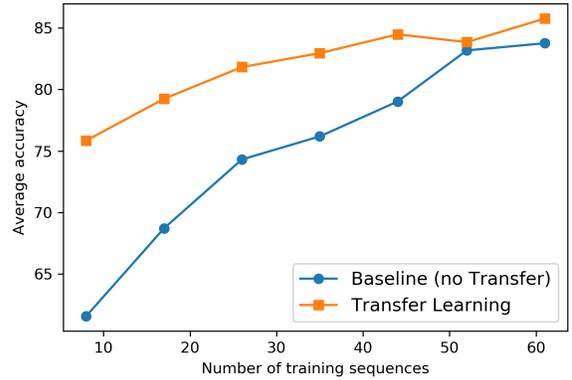


Fig. 6: Impact of increasing size of target dataset. The slow increase in accuracy of the transfer learning suggests that the network carries significant momentum from the weights trained using the peg transfer task.

network was retrained on this training set starting from the peg-transfer weights.

The impact of having more training data in the target domain is presented in Figure 6. The comparison is against the baseline of no-transfer learning. As shown, even with as little as 8 training sequences the classification accuracy of the transfer learning setup is 76% which is 22.6% better than the baseline. This suggests that the amount of training data required in the new surgical domain is reduced using a transfer learning setup.

2) *Use of each layer in classification:* For object recognition networks, transfer learning is usually achieved by re-training the last few layers [35]. This is because the object specific discriminating features are learnt in the last few layers, whereas the initial layers produce low level image features such as edges, shapes, or textures. To analyze the proposed network similarly, the effect of retraining the different layers is evaluated. Our model comprises of an LSTM layer which can infer using the hidden states based on the history. Following this is a series of fully connected dense layers which use the outputs of LSTM layer to generate features for classification.

Figure 7 presents the impact of re-training the different

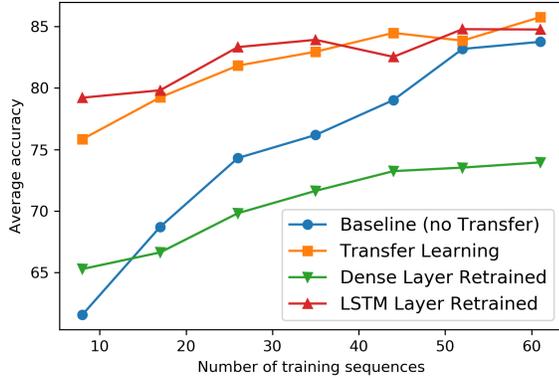


Fig. 7: Impact of re-training specific layers for transfer learning. Re-training dense layer only performs worse than the baseline, which suggests that most of the surgeme specific knowledge is capture by the LSTM layer.

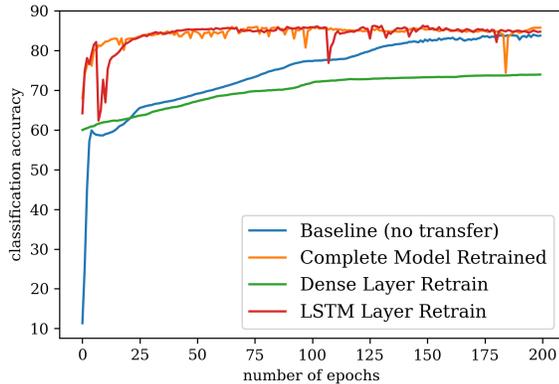


Fig. 8: Impact of transfer learning on the convergence of the neural network. The transfer learning setup converges after about 50 training steps as compared to the non-transfer learning setup which requires about 3 times more training steps.

layers of the classification network. We observe that the re-training only LSTM layer achieves the same effect as re-training the entire network. This implies that the LSTM layer captures the domain-specific discriminating features required for surgeme classification. Further, retraining only the dense layers performs worse than the baseline. This signifies that re-training the LSTM layer is essential for transfer learning across sequential data such as surgemes.

3) *Learning Rate of the transfer learning*: All the networks are trained using stochastic gradient descent optimizer with a constant learning rate of 0.005. The transfer learning allows the model to train faster as compared to the baseline without transfer learning. To establish this fact, we compared the classification accuracy after each training step. We present the result for the networks are trained with 61 training sequences.

Figure 8 shows the effect of transfer learning on the

convergence rates. The networks initialized with pre-trained weights already obtain an initial push to move towards the optima resulting in a faster convergence rate. The analysis of convergence rates shows that the transfer learning setup can not only provide advantages when there is little data in the target domain but also when the training time is limited. This is specially useful in cases where an emergency situation may allow only few hours for training the robot before deployment.

VII. CONCLUSIONS

In austere environments, such as battlefields, even performing the debridement task using a tele-surgical robotic systems can be life saving in the absence of a medic. To eliminate severe impact of limited or no connectivity on the implementation of tele-surgery robotic systems execution using surgemes is considered. In this paper we considered developing a classifier for surgeme recognition for semi-autonomous performance of the debridement task. We developed a simulator where a Taurus robot performs debridement task. Kinematic data for debridement task was collected and used to train an LSTM classifier. Such a classifier is then used to infer the current surgeme being performed. We showed that proposed LSTM framework can classify surgemes with an accuracy of 84.4%.

To reduce the data requirements and allow faster deployment, transfer learning is used. Data collected from peg transfer task was used to pre-train the network for transfer learning. By transferring the weights from surgeme recognition on the peg transfer task, the performance of the classifier increases by 10% when only 50% of the training data is available. It was shown that the non-transfer learning setup requires 50% more data to achieve the performance of the transfer learning setup. Ablation studies were conducted to understand the role of each layer in the transfer learning task. It was found that it is necessary to retrain the LSTM layer as the temporal structure of the kinematic sequences vary between the peg transfer task and the debridement task. Further, we found that the transfer learning models converges to the solution faster than the non-transfer learning method. We conclude that the proposed transfer learning allows the surgical system to be easily transferable to other surgical tasks.

ACKNOWLEDGMENTS

This work was supported by the Office of the Assistant Secretary of Defense for Health Affairs under Award No. W81XWH-18-1-0769 and by the NSF RoSe-HUB Center under CNS-1439717. Opinions, interpretations, conclusions, and recommendations are those of the authors and are not necessarily endorsed by the funders.

REFERENCES

- [1] P. Garcia, J. Rosen, C. Kapoor, M. Noakes, G. Elbert, M. Treat, T. Ganous, M. Hanson, J. Manak, C. Hasser, *et al.*, "Trauma pod: a semi-automated telerobotic surgical system," *The International Journal of Medical Robotics and Computer Assisted Surgery*, vol. 5, no. 2, pp. 136–146, 2009.

- [2] T. Haidegger, "Autonomy for surgical robots: Concepts and paradigms," *IEEE Transactions on Medical Robotics and Bionics*, vol. 1, no. 2, pp. 65–76, 2019.
- [3] J. Loyall, M. Gillen, J. Cleveland, K. Usbeck, J. Sterling, R. Newkirk, and R. Kohler, "Information ubiquity in austere locations," *Procedia Computer Science*, vol. 10, pp. 170–178, 2012.
- [4] H. C. Lin, I. Shafran, D. Yuh, and G. D. Hager, "Towards automatic skill evaluation: Detection and segmentation of robot-assisted surgical motions," *Computer Aided Surgery*, vol. 11, pp. 220–230, Jan. 2006.
- [5] S. H. Y. John M. Fossaceca, "Artificial intelligence and machine learning for future army applications," 2018.
- [6] Y. Gao, S. S. Vedula, C. E. Reiley, N. Ahmidi, B. Varadarajan, H. C. Lin, L. Tao, L. Zappella, B. Béjar, and D. D. Yuh, "JHU-ISI gesture and skill assessment working set (JIGSAWS): A surgical activity dataset for human motion modeling," in *MICCAI Workshop: M2CAI*, vol. 3, p. 3, 2014.
- [7] N. Madapana, M. M. Rahman, N. Sanchez-Tamayo, M. V. Balakuntala, G. Gonzalez, J. P. Bindu, L. N. V. Venkatesh, X. Zhang, J. B. Noguera, T. Low, R. Voyles, Y. Xue, and J. Wachs, "Desk: A robotic activity dataset for dexterous surgical skills transfer to medical robots," in *2019 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS-2019)*, 2019.
- [8] R. DiPietro, C. Lea, A. Malpani, N. Ahmidi, S. S. Vedula, G. I. Lee, M. R. Lee, and G. D. Hager, "Recognizing Surgical Activities with Recurrent Neural Networks," in *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2016* (S. Ourselin, L. Joskowicz, M. R. Sabuncu, G. Unal, and W. Wells, eds.), Lecture Notes in Computer Science, pp. 551–558, Springer International Publishing, 2016.
- [9] G. P. Moustris, S. C. Hiridis, K. M. Deliparaschos, and K. M. Konstantinidis, "Evolution of autonomous and semi-autonomous robotic surgical systems: a review of the literature," *The International Journal of Medical Robotics and Computer Assisted Surgery*, vol. 7, no. 4, pp. 375–392, 2011.
- [10] C. E. Reiley and G. D. Hager, "Task versus Subtask Surgical Skill Evaluation of Robotic Minimally Invasive Surgery," in *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2009* (G.-Z. Yang, D. Hawkes, D. Rueckert, A. Noble, and C. Taylor, eds.), Lecture Notes in Computer Science, pp. 435–442, Springer Berlin Heidelberg, 2009.
- [11] L. Tao, E. Elhamifar, S. Khudanpur, G. D. Hager, and R. Vidal, "Sparse Hidden Markov Models for Surgical Gesture Classification and Skill Evaluation," in *Information Processing in Computer-Assisted Interventions* (P. Abolmaesumi, L. Joskowicz, N. Navab, and P. Jannin, eds.), Lecture Notes in Computer Science, pp. 167–177, Springer Berlin Heidelberg, 2012.
- [12] L. Tao, L. Zappella, G. D. Hager, and R. Vidal, "Surgical Gesture Segmentation and Recognition," in *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2013* (K. Mori, I. Sakuma, Y. Sato, C. Barillot, and N. Navab, eds.), Lecture Notes in Computer Science, pp. 339–346, Springer Berlin Heidelberg, 2013.
- [13] N. Ahmidi, L. Tao, S. Sefati, Y. Gao, C. Lea, B. B. Haro, L. Zappella, S. Khudanpur, R. Vidal, and G. D. Hager, "A Dataset and Benchmarks for Segmentation and Recognition of Gestures in Robotic Surgery," *IEEE Transactions on Biomedical Engineering*, vol. 64, pp. 2025–2041, Sept. 2017.
- [14] C. Lea, G. D. Hager, and R. Vidal, "An Improved Model for Segmentation and Recognition of Fine-Grained Activities with Application to Surgical Training Tasks," in *2015 IEEE Winter Conference on Applications of Computer Vision*, pp. 1123–1129, Jan. 2015.
- [15] C. Lea, R. Vidal, and G. D. Hager, "Learning convolutional action primitives for fine-grained action recognition," in *2016 IEEE International Conference on Robotics and Automation (ICRA)*, pp. 1642–1649, May 2016.
- [16] M. M. Rahman, N. Sanchez-Tamayo, G. Gonzalez, M. Agarwal, V. Aggarwal, R. M. Voyles, Y. Xue, and J. Wachs, "Transferring dexterous surgical skill knowledge between robots for semi-autonomous teleoperation," in *28th IEEE International Conference on Robot and Human Interactive Communication (Ro-Man-2019)*, 2019.
- [17] M. M. Rahman, M. V. Balakuntala, G. Gonzalez, M. Agarwal, U. Kaur, V. L. N. Venkatesh, N. Sanchez-Tamayo, Y. Xue, R. M. Voyles, V. Aggarwal, and J. Wachs, "Sartres: a semi-autonomous robot teleoperation environment for surgery," *Computer Methods in Biomechanics and Biomedical Engineering: Imaging & Visualization*, vol. 0, no. 0, pp. 1–8, 2020.
- [18] G. T. Gonzalez, U. Kaur, M. Rahman, V. Venkatesh, N. Sanchez, G. Hager, Y. Xue, R. Voyles, and J. Wachs, "From the dexterous surgical skill to the battlefield—a robotics exploratory study," *Military Medicine*, vol. 186, no. Supplement.1, pp. 288–294, 2021.
- [19] M. M. Rahman, R. M. Voyles, J. Wachs, and Y. Xue, "Sequential prediction with logic constraints for surgical robotic activity recognition," in *30th IEEE International Conference on Robot and Human Interactive Communication (Ro-Man-2021)*, 2021.
- [20] C. Tan, F. Sun, T. Kong, W. Zhang, C. Yang, and C. Liu, "A survey on deep transfer learning," in *International conference on artificial neural networks*, pp. 270–279, Springer, 2018.
- [21] D. Mahajan, R. Girshick, V. Ramanathan, K. He, M. Paluri, Y. Li, A. Bharambe, and L. Van Der Maaten, "Exploring the limits of weakly supervised pretraining," in *Proceedings of the European Conference on Computer Vision (ECCV)*, pp. 181–196, 2018.
- [22] X. Wang, Y. Peng, L. Lu, Z. Lu, M. Bagheri, and R. M. Summers, "Chestx-ray8: Hospital-scale chest x-ray database and benchmarks on weakly-supervised classification and localization of common thorax diseases," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 2097–2106, 2017.
- [23] B. Neyshabur, H. Sedghi, and C. Zhang, "What is being transferred in transfer learning?," in *Advances in Neural Information Processing Systems* (H. Larochelle, M. Ranzato, R. Hadsell, M. F. Balcan, and H. Lin, eds.), vol. 33, pp. 512–523, Curran Associates, Inc., 2020.
- [24] F. Zhuang, Z. Qi, K. Duan, D. Xi, Y. Zhu, H. Zhu, H. Xiong, and Q. He, "A comprehensive survey on transfer learning," *Proceedings of the IEEE*, vol. 109, no. 1, pp. 43–76, 2020.
- [25] M. Oquab, L. Bottou, I. Laptev, and J. Sivic, "Learning and transferring mid-level image representations using convolutional neural networks," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 1717–1724, 2014.
- [26] J.-T. Huang, J. Li, D. Yu, L. Deng, and Y. Gong, "Cross-language knowledge transfer using multilingual deep neural network with shared hidden layers," in *2013 IEEE International Conference on Acoustics, Speech and Signal Processing*, pp. 7304–7308, IEEE, 2013.
- [27] A. R. Zamir, A. Sax, W. Shen, L. J. Guibas, J. Malik, and S. Savarese, "Taskonomy: Disentangling task transfer learning," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 3712–3722, 2018.
- [28] D. Zhang, Z. Wu, J. Chen, A. Gao, X. Chen, P. Li, Z. Wang, G. Yang, B. Lo, and G. Yang, "Automatic microsurgical skill assessment based on cross-domain transfer learning," *IEEE Robotics and Automation Letters*, vol. 5, no. 3, pp. 4148–4155, 2020.
- [29] Y.-Y. Tsai, B. Huang, Y. Guo, and G.-Z. Yang, "Transfer learning for surgical task segmentation," in *2019 International Conference on Robotics and Automation (ICRA)*, IEEE, May 2019.
- [30] K. A. Nichols, A. Murali, S. Sen, K. Goldberg, and A. M. Okamura, "Models of human-centered automation in a debriement task," in *2015 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pp. 5784–5789, IEEE, 2015.
- [31] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural computation*, vol. 9, no. 8, pp. 1735–1780, 1997.
- [32] G. Gonzalez, M. Agarwal, M. V. Balakuntala, M. M. Rahman, U. Kaur, R. M. Voyles, V. Aggarwal, Y. Xue, and J. Wachs, "Deserts: Delay-tolerant semi-autonomous robot teleoperation for surgery," in *2021 IEEE International Conference on Robotics and Automation (ICRA)*, IEEE, 2021.
- [33] S. Ji, W. Xu, M. Yang, and K. Yu, "3d convolutional neural networks for human action recognition," *IEEE transactions on pattern analysis and machine intelligence*, vol. 35, no. 1, pp. 221–231, 2012.
- [34] M. Müller, "Dynamic time warping," *Information retrieval for music and motion*, pp. 69–84, 2007.
- [35] Z. Yang, W. Yu, P. Liang, H. Guo, L. Xia, F. Zhang, Y. Ma, and J. Ma, "Deep transfer learning for military object recognition under small training set condition," *Neural Computing and Applications*, vol. 31, no. 10, pp. 6469–6478, 2019.