Deep learning in medical Robotics application: past, now and future

1st Redian Copa
Robotics and Embedded Systems
Technische Universität München
Munich, Germany
redian.copa@tum.de

2nd Stefan Dünhuber
Robotics and Embedded Systems
Technische Universität München
Munich, Germany
s.duenhuber@tum.de

3rd Ixhen Hasani
Robotics and Embedded Systems
Technische Universität München
Munich, Germany
ixhen.hasani@tum.de

4th Carolin Hecking-Veltman
Robotics and Embedded Systems
Technische Universität München
Munich, Germany
carolin.hecking@tum.de

5th Jennifer Mock
Robotics and Embedded Systems
Technische Universität München
Munich, Germany
jennifer.mock@tum.de

6th Nhan Tran
Robotics and Embedded Systems
Technische Universität München
Munich, Germany
nhan.tran@tum.de

7th Liudongnan Yang
Robotics and Embedded Systems
Technische Universität München
Munich, Germany
liudongnan.yang@tum.de

***Abstract*—In the past years great progress has been made in the area of Machine Learning using Deep Neural Networks. These networks have enabled advances particularly in computer vision and natural language processing. Deep Learning has been applied to different problems in the medical field such as classifying tissues in images or segmenting tumors. Another interesting possible application of Deep Learning is Medical Robotics. Robots are used to perform surgeries, support doctors in medical procedures and additionally are used to enhance prosthetics. In this work, the advances that have been made in applying deep learning to medical robotics are presented. We will firstly introduce a brief history about Deep Neural Networks and the principals of Deep Neural Networks. Furthermore three promising areas of applications in medical robotics will demonstrate: Firstly, improving the processing of electromyography signals that can control movements of prosthetics. Secondly, using Convolutional Neural Networks for image processing before and during surgery. Lastly, using Deep Learning for improving surgical skill assessment and automating subtasks like tie knots in robot-assisted laparoscopically surgery. In all three application areas the past developments, state of the art and future approaches to use Deep Learning in Medical Robotics are given.**

***Index Terms*—deep learning; convolutional neural network (CNN); recurrent neural network (RNN); robotic-assisted surgery; image analysis**

I. A Brief Introduction of Deep Learning

A brief history of deep learning

Linear regression was first used by Gauss in the early 19th Century. After that, some of the principles of regression slowly transformed into Deep Learning Algorithms, which have evolved to current trends in robotic systems. In past decades, scientists were inspired by biological brain cells and started further developing the non-linear model using stacked layers in order to create powerful models. These are call multi-layer perceptrons (MLP). In the 1960s, a few researchers independently figured out how to differentiate MLPs and by the 1980s, this resulted in a popular method for training them, called backpropagation (BP). It could be shown that MLPs as a result were universal function approximators, meaning they can fit any data by using hidden units. Neural networks were successfully applied to robotics controlling systems in the early 1980s. In 1989, Pomerleaus ALVINN formally showed that neural networks were effective for helping vehicles to drive within a lane.

In the 2000s, as graphical techniques developed, researchers tried using graphical processing units (GPU) to implement artificial neural networks as well as striving to get a better performance. In 2006, famous Canadian computer scientist Hinton developed an efficient training procedure for multi-layer neural networks. This was the starting point for the acceleration of progress in DL that can currently be observed. In 2011 TNLDR showed that NNs can be used to indicate status as well as parameters in unsupervised training on images generated by a simulated robot. In the present, artificial intelligent systems are being applied to many new tasks. Over and over again, the great capability of performing these tasks such as visual cognition at human-like performance is shown.

Deep learning has been making many contributions to advanced image recognition techniques, and now shows its powerful ability in recognition related area. [1]

Introduction

Deep learning is about training large artificial neural networks. Professor Andrew Ng, the former Chief Scientist at Baidu, who founded Google Brain that funded production of deep learning techs. He has once described in a Speech about the idea of deep learning that it makes learning algorithms much better and easier to use. Furthermore it makes revolutionary advances in machine learning and AI. He also mentions that he believes this is our best shot at progress towards real AI. According to Andrew the core of deep learning is that we now have the computer power and the accessible data to train neural networks. What is also worth mentioning is that the more data network gets trained the higher the Performance will be. This is quite different to other machine learning techniques that reach a plateau performance when the dataset is large.
[2] Because Deep Neural Network(DNN) is a common method of multi-layered nonlinear dynamics, they form a compact representation from raw data in robotic systems.

classification of learning

Supervised learning: Supervised learning is the machine learning method which input is already marked with labels on all training data sets. In supervised learning, each entity represents an input object (usually in the form of a vector) and expects an output value. The learning method infers a result which will then be mapped to an input entity.

Unsupervised learning: In contrast to supervised learning, unsupervised learning has the input data set without a label, the machine tends to gather information in a more active way, the typical task of unsupervised learning contributes clustering abnormal detection.

Reinforcement learning: In many real-world problems,it is very difficult to establish the solution, but it is easy to decide whether a task was achieved successfully. Therefore Reinforcement learning was introduced where an agent interacts with its environment to maximize an external reward signal to learn which action to choose in a certain situation.

II. Structures of Neural Networks

Common Structures

As the structure of the neural network in figure 1.A. suggests, the model will be trained by a set of data $< x_1, y_1 >$, $< x_2, y_2 >, ..., < x_n, y_n >$. Optimization is applied to minimize the loss. The regression loss will be measured in sum-squared: $\sum_i^n (y_i - \hat{y}_i)$. And for classification problems often measured with cross-entropy: $-\sum_i^n (y_i - \hat{y}_i)^2$. In figure 1.B. is a common model of an auto-encoder that promotes unsupervised learning. It requires 2 DNNs: an encoder and a decoder, x is given by the user, s is the hidden part or "latent" that is generated by DNN encoder. Generative models have a close relation to auto-encoders, which utilize the decoder in order to anticipate or predict observations from the internal state. figure 1.C. is a type of recurrent neural network (RNN), which is often trained with an approach called back-propagation through time, that allows the model to have a few features such as long & short-term memory units. The configuration specifies that: u is control signal that contains recent observations, s is an internal representation of the future state, x is an output vector that represents anticipated future observations. figure 1.D. models control policy, which is based on reinforcement learning. The variable q evaluates the potential control vector, s refers to the state and u is the control vector. Despite the tremendous number of neural network models, the 4 models mentioned above are the basic and the most common models that were applied firstly in robotic surgery. [1]

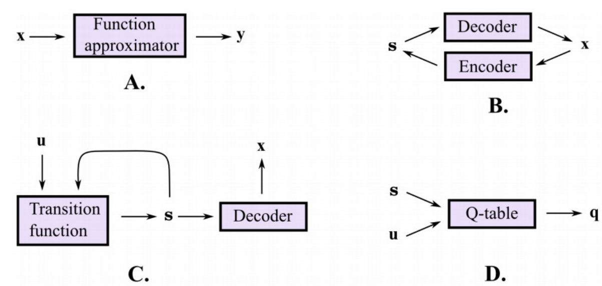

Fig. 1. structure of common model including different types. [1]

Convolutional Neural Networks (CNN)

A CNN is commonly composed of three parts. The first part is the input layer, the second part consists of a combination of n convolutional layers and pooling layers. The third part is a multi-layer perceptron classifier. Furthermore, there are 3 important definition to be mentioned: 1. **Local receptive fields.** In contrary to fully connected layers, convolutional layers do not have access to the entire input but instead are limited to field of the input.
2. **Shared weights.** shared weights,for each neuron of the hidden layer,uses the same weight as well as bias i.e.if we use 3*3 receptive field for hidden neoron at position i,j $\sigma(b + \sum_{l=0}^{2} \sum_{m=0}^{2} w_{l,m} a_{i+l,j+m}))$, σ stands for heuristic function(or RELU), b means shared bias ,w is the input array,and $a_{x,y}$ representation input value of the position x,y.
3. **Pooling.** Another crucial part of the CNN structure are pooling layers. Specifically, pooling layers obtain the feature map and compute a concentrated feature. Finally, the CNN consists of many convolutional layers, such that they generate several feature maps. This is visualized in figure 2. This structure contributes to the reduction of parameters of the layers.

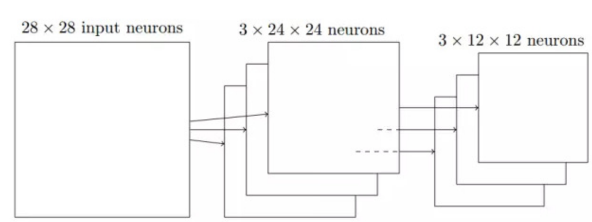

Fig. 2. structure of Convolutional layers with pooling phase.

Convolutional Layer: Unlike traditionally used fully-connected layers, convolutional layers use the same weights to operate all across the input space. This can help to substantially reduce the number of weights. This in turn benefits improving the performance of processes that have millions of weights as in image processing. [2]

III. Tracking of medical Instruments during Surgery

Medical Images are getting a more important role in helping to identify abnormalities and/or diseases. The most prominent way of analyzing images is Deep Learning (DL) with the most used method Convolutional Neural Network (CNN) which is used for 2D-images as well as 3D-images. In the following sections we are going to take a deeper look at image Segmentation with a few prominent examples of Segmentation methods and the field of application surgical tool detection.

A. Segmentation

Image machine learning research makes a distinction in four types of analysis.

- Localization - fixes the position of an object in an image and draws a box around it
- Detection - draws a box around multiple objects that don't have to be of the same class
- Segmentation - draws not just a box but outlines one or multiple objects and labels them
- Registration - fitting one image onto another

[3] In the following passage we are going to focus on Segmentation methods. In the past anatomical segmentation was done by hand. Which meant going through a stack of images one by one and putting a lot of time into this. Segmentation research has been mostly on 2D images although 3D research has had some success for example in the DeepMedic CNN. A problem both image types share and which further hinders teaching DL algorithms is the lack of available data, since medical images are rarely public available on a grander scale. Even less frequent are labeled images.

1) Thresholding: Segmentation method in which a binary image of the original image is constructed to reduce the complexity of the image, so that the process of classification and recognition is simplified. It splits the image in smaller segments to define the boundary in the image based on intensity levels. Usually the image is transformed in a gray-scale or one color spectrum. [5]

2) Edge based: One of the two main Segmentation methods with the other being Region based Segmentation. It's input is a gray-scale level image to find discontinuities in them. Edges are local changes in the image intensity, that typically occur on the boundary between two regions. This makes higher level image analysis possible. [6]

3) Region based: These classify the image into subregions. The subregions are defined by predefined rules. For example the color or gray-scale has to be the same . Region growing is a method which group pixels in the image into subregions or regions based on predefined rules. Region splitting on the other hand divides the image into subregions. [6]

Fig. 3. Segmented picture with low-level Thresholding Segmentation [4]

B. General approach and problems

In Medicine it is becoming more important to analyze Videos of the surgery process in real-time. The reason being that minimal invasive surgery(MIS) is obviously more lucrative for clinic and patients. Less painful, lesser blood loss, recovery time etc. The drawback for the execution and the surgeons are that they have to be a lot more precise and accurate in their hand movements on top of the operations lesser visibility. So that more information through automatic Video/Image analysis is welcome and necessarily. Another use for this is fully robotic surgery which has as an example the da Vinci robots or Raven. [7] Segmentation of surgical tools in images is difficult because of visual obstacles like lightning challenges such as light from reflections or shadows, medical challenges such as blood, smoke and the general complexity of background textures and changing conditions as body fluids, motion blur and the changing background because of moving organs. This leads to organs and instruments looking more similar on pictures then they should even though they are differently colored. [8] Early ideas to handle this problem included modifying the instruments to make them more visible with various markers. For example optical trackers, acoustic trackers, color/shape markers and so on. But this complicates clinical tasks as it is not practical. [9] A further problem comes into play when looking at real-time segmentation because the time the image segmentation needs to be executed becomes important so that the result has to be present for the surgeon operating.

C. Research

Most research has been done on laparoscopic and endoscopic MIS with second most researched retinal micro-surgeries while the test matter was mostly human bodies porcine bodies were also used. 2D images were more represented then 3D images consisting of 2D images to get more information of depth of field. The type of research ranges from clinical realistic but no control over environment to clinical unrealistic in a controlled setting. The data's quantity on which the research is done can be divided into experiments with only one video sequence and experiments with more than five video sequences which are more recent. [7] As a consequence the image quantity ranges from a little pile of less then a hundred to a big set of more then thousand pictures. The type of tool that is segmented is often not specifically mentioned but most segmentation and tracking has been done on videos with two instruments although more or less are also featured.

D. Methods

The most often used machine learning methods can be grouped in four essential approaches. [7]

1) Color: The most basic and researched method which detects instruments based on their color. Different color spectra can be used. RGB is one of them and HSV/HSL another one which is more complex and allows a better separation of colors because saturation and lightness come into play. CIE L*a*b is also used but comes with the drawback of needing more Bit.

2) Gradient: This method often uses color segmented images as an input and tries to detect the rise and fall of edges to determine were the tool is and in which relation to the tissue it is. [7] Most commonly this is done by looking at a specific color component or intensity. An option to strengthen this method is to use a texture method before so that the areas of the image are better defined.

Fig. 4. Segmented picture with saturation adjustment [7]

3) Texture: Segments the image by keeping parts together that share the same or a similar pattern. Using SIFT features is a promising method that has already seen some success although it has not yet been used often in the context of MIS.

4) Shape: The least used [7] of the four presented methods it uses region-based Segmentation to acquire the information which basic shapes like circles, square, rectangle and so on are present in the given picture.

E. Real-time Tracking and Detection

Most research looks at the image analysis as a binary segmentation problem but for a more precise outcome it would be important to segment the different parts of the surgical tool. Efficient algorithms include region-based CNNs and Region Proposal Networks who achieved near real-time detection with 17fps. For laporoscopic videos EndoNet has been put forward as an solution for tool presence detection. Instead of a state of the art Convolutional neural Network modified CNNs in shape of Fully Convolutional Networks, who are better fitted to semantic labeling then CNNs who specify in detection, were also able to get good test results. [10] [9]

Tracking is used to detect small motions between two images in at its best in real time. The FCN segments each individual image and then registers the last segmented frame with the current one. When the segmentation is finished by the FCN it updates the last segmentation mask and stores it in synchronized memory. Should the instruments move to fast the FCN is not able to detect and segment the picture.

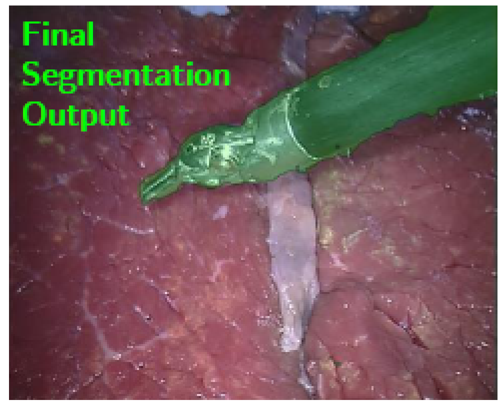

Fig. 5. An example for a segmented picture with FCN [9]

IV. Image Processing using Deep Learning in Surgical Applications

Image processing including segmentation, labeling und recognition is one of the key applications Deep Learning (DL) has been used for so far. The most commonly used Neural Network architecture chosen for image input data is the convolutional neural network (CNN). This allows for translation-independent processing of the images, where a feature may be found anywhere in the image to be detected classified. There are many possible applications in the medical field for DL. Those that are used related to surgery are presented in the following. The main questions that will be answered are:

- What progress has been made in the past?
- What applications are currently being worked on?
- What are further possible applications in the future?

A. Segmentation and Labeling of Abnormal Tissue

As stated in [11], CNNs can be used for image segmentation in cancer patients. An example result is shown in 6. The image is fed into a CNN which then labels the image pixel-wise into tumor and healthy tissue. The architecture used for this problem most commonly is the U-Net. This enables doctors to effectively locate the tumor and make informed choices about treatment. Another important task of the segmentation is monitoring tumor growth ([12]). Furthermore, the automatic segmentation of the tumor helps removing tissue during surgery with significantly increased precision. Without the

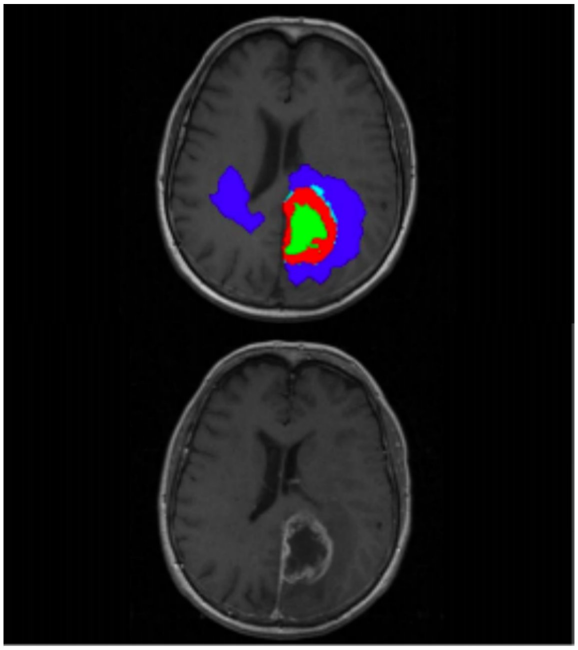

Fig. 6. Segmented Glioblastoma. Source: [13]

(semi-)automatic segmentation by DL using CNNs, the tumor has to be segmented in the images manually which is not feasible due to time intensity and economic reasons. This is worsened by the fact that 3D medical imaging is realized with up to 250 2D images visualizing different slices of the brain or other body parts. The CNN is trained using a training set of manually labeled medical images, e.g. CT images or MRI images. Such methods for tumor segmentation are benchmarked using BRATS (brain tumor image segmentation benchmark). [11] reports precisions produced by CNN models that reaches that of humans. Recent research has been directed towards segmenting the tumor into subregions as stated in [12]. In this work DenseNet is used which is a CNN architecture that was originally developed for image classification problems.

B. Location of Surgery Site

Using this methodology and intraoperational MRI ([14]), DL can also help surgeons locating and identifying malignant tissue during surgery. So far this has been used predominantly in resection surgeries for glioblastoma, a kind of brain tumor. It is done using intraoperational MRI (iMRI). The use of iMRI has already significantly improved the share of a tumor that is removed in a surgery ([14]). In the future, image segmentation by DL during surgery may enhance this method by clearly indicating the location of a tumor in real-time. Challenges for this approach may include the real time component since MRI is time intensive.

C. Image Recognition in Endoscopic camera images

In the past decades a shift from open surgery to endoscopic surgery has taken place. The advantages of endoscopic surgery include faster recovery of the patient. One challenge in endoscopic surgery is posed by the limited field of view. In [15], multiple techniques that work towards solving this problem are presented and the work shows that this is a relevant field of research. At the time that work was published, DL had not made many advances yet. The subproblems that are mentioned in [15] include segmentation, 'Implicit Non-rigid Shape and Motion Recovery' and tracking of objects. DL is a new approach how these challenges may be solved.

In [16], shot classification that performs semantic segmentation of images obtained during endoscopic surgery is done using DL. Semantic segmentation segments an image and labels the different parts as objects. In [16] the CNN architectures AlexNet and GoogLeNet were used. While the semantic segmentation was so far only used for medical training, it is well conceivable that semantic segmentation will be used during surgery in the future. Semantic segmentation can be reused to estimate the position of the camera in relation to the non-rigid patient which may help surgeons to perform the surgery with higher precision. Another possible application may be heart surgery in the far future where the instrument automatically moves such that it holds constant position relative to the heart.

D. Virtual surgery

Virtual surgery uses robotic arms and visual feed back to the surgeon. It enables higher precision in surgeries. Just as in non-virtual endoscopic surgery, additional information about the presented images can be very helpful to increase quality and precision of surgery. In the future, telerobotic surgery ([17]) might be enhanced by generating information about the surgery from images of the rest of the body or vital signs. By feeding this information into a NN that was trained on data from many previous surgeries, a surgeon that is located at a different place and interacts with the patient via robotic surgical arms can be given condensed important information about the state of the patient. By additionally encoding patient history or e.g. other current injuries and diseases this system could be enhanced.

E. Documentation via Image Recognition

As stated in [16], the images recorded during endoscopic surgery are an important source for documentation in hospitals. Using DL this may be extended to a point where large parts of a surgery documentation are automatically generated from the recorded images using DL. A starting point for such a solution would be semantic segmentation of the image time series. Subsequently, the matching of image subseries to text description would have to be learned. Next to CNNs, this will likely additionally require recurrent neural networks (RNN) to analyze the images in the time dimension. A possible training set for this problem could be recordings from endoscopic cameras of past surgeries with the corresponding text documentation.

In addition, having access to a large data base of documented other surgeries, this approach could be extended to predict complications that are likely to happen at the current surgery step in real-time. While surgeons are obviously trained to keep possible complications in mind at any time, such an approach can draw information from far more surgeries. This could evolve into an intelligent assistent that understands the current status of the surgery and e.g. autonomously hands the right instruments to the surgeon supporting her in time critical situations.

F. Natural Language Processing for Documentation and Recommendations of Surgical Steps

Another field that is currently being revolutionized by DL is Natural Language Processing (NLP). It is mainly performed using RNNs. There are many different problems that are approached by NLP such as translating from one language to another, understanding the context of words or speech recognition.

In surgical applications, NLP might be used for documentation in the future. This could be done by recording speech in the operating room and performing speech recognition. There obviously is quite a concern about privacy related to this, that would have to be addressed carefully. Since documentation is a very time intensive process this would give the hospital staff more time to spend on patient care. The approach would require surgeons to describe the surgical steps during the surgery to work optimally. When being used in combination with endoscopic camera images this approach could again warn the surgeons of possible complications. Furthermore, such a system could even recommend next steps when trained on a large data base of past surgeries.

V. Deep Learning for Robot-assisted surgery

One main area for medical robotics application is the field of Robot-assisted surgery. The term defines technological developments that use robotic systems to aid in surgical procedures. In the area of deep learning for Robot-Assisted surgery two main applications must be mentioned:

1) Improving Surgical Skill Assessment and Task Recognition
2) Automating subtasks like tie knots in robot-assisted laparoscopically surgery

A. Improving Surgical Skill Assessment and Task Recognition in Robot-assisted Surgery with Deep Neural Networks

1) Current Surgical Skill Assessment: Skill level of surgeons is an important factor in surgery and can affect clinical outcomes [18]. However, there are many different methods of teaching. Current approaches of objective skill assessment techniques can be divided into three main categories [19]. The first category is structured human grading. Skill level varies among surgeons and can improve with training, teaching and experience. Several factors, which include jugdment and decision-making, manual dexterity and cognitive capabilities determine the skill level of a surgeon. Those factors can be taught in an apprentice-style of teaching. However, this approach is time consuming and requires human resources. Structured human grading is an observational approach. Surgeons can provide verbal feedback while observing a student.

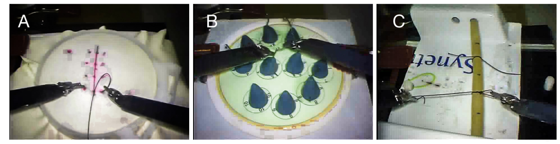

Fig. 7. Robot-assisted minimally invasive training. Three different tasks: (A) Suturing, (B) Needle-passing, (C) Knot-tying

This grading technique uses rated checklists to standardize the evaluation. [19] The second category is descriptive statistic analysis. These methods compute features from motion observations to describe skill levels. Movement time, path length or curvature are different variables, which gives information about the skill level of a surgeon. This approach requires task-specific knowledge. Furthermore, best metrics has to be defined and also be generalized enough to apply to different surgery types. [20] The third category is predictive modeling-based methods. This approach predicts surgical skills from motion data. With descriptive modeling, raw motion are transformed to interpretations and summary features. Machine learning algorithms are commonly used for modeling, such as k-nearest neighbors (kNN). Although the descriptive modeling-based seems like a valid approach by revealing skill patterns and underlying operation structures, features of complex surgical motion can be discarded within the feature extraction and selection process. By using a generative modeling algorithm such as Hidden Markov Model (HMM), skill models were trained and can achieve accuracy ranging from 94.4 percent to 100 percent. However, this approach requires a large amount of time. [18], [20]

2) Improving Surgical Skill Assessment: Technical training programs ensure trainees to develop surgical skills to perform difficult operations. Teaching techniques for surgical skill assessment can be time consuming. For better quality and safety in surgery, trainees have to achieve a required skill level before operating on patients. Most assessments of skills are still performed with rating scales and outcome-based analysis. These methods require expert monitoring and manual ratings, which can be inconsistent, because of different human interpretations. [19] To increase the efficiency and effectiveness of assessment, these conventional methods are no longer adequate in advanced surgery. Modern robot-assisted surgical systems can collect a high volume of data with their surgical robots or simulators. Deep learning is an approach to increase efficiency and effectiveness in surgical skill assessment. [18]

3) Robot-assisted Surgery with Deep Neural Network: Deep structured learning contains a set of learning methods that allow machines to process and learn from input data. Through several self-learning steps, these algorithms can progressively discover abstract representations. Feature learning is automatically employed without manual extraction and selection. Simulations can can ensure a safe learning environment for the trainee. [18]

4) SATR-DL: The surgery motion analytic framework with a multi-output deep learning architecture, SATR-DL, achieves

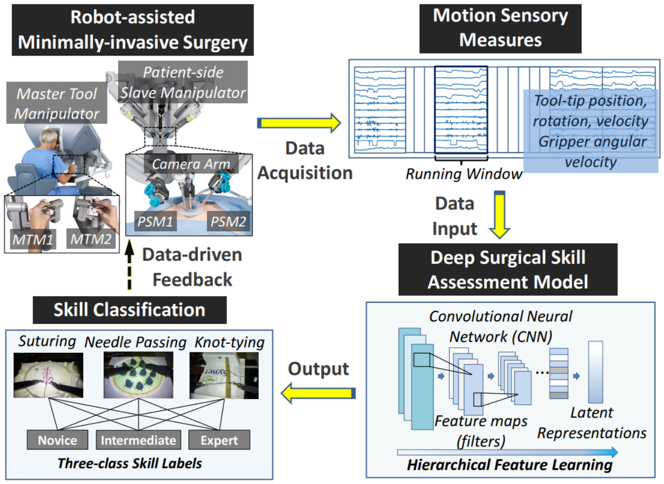

Fig. 8. End-to-end framework for skill assessment. Source: [18]

a trainee skill analysis and task recognition by integrating a Convolutional Neural Network (CNN) and Gated Recurrent Unit (GRU) network. The assessment of surgical skills can be formalized as a supervised three-class classification problem. Surgical robot end-effectors measure motion kinematics. [19] The input is a multivariate time series (MTS) of these measured motion kinematics and the output is the predicted labels, which represents the skill level of trainees.This end-to-end framework uses motion sensory measures like tool-tip position, rotation and angular velocity as data input to a deep surgical skill assessment model. [18]

The deep architecture contains a 10-layer convolutional neural network. The network length is chosen after trial-and-error from training and validation process. [18] [19] The 10 layers are distributed in 5 different types of layer. First the input data is processed by three convolution-pooling stages, which contains a convolutional and a max-pooling layer. The output of the convolution layer is taken as an input for the max-pooling operation. This step downsamples the extracted feature map. Afterwards, these data are processed in the fully-connexted layers and apply a softmax logistic regression. [18], [20]

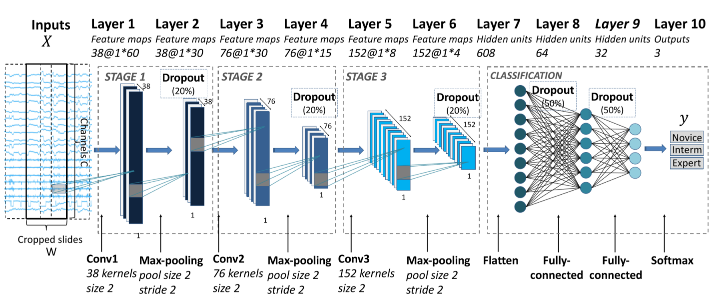

Fig. 9. Architecture of 10-layer convolutional neural network. Source: [18]

5) Outlook of skill Evaluation in robot-assisted surgery with Machine Learning: By using supervised machine learning, robotic-assisted surgical tasks can be distinguish between expert and novice performance. Future research could emphasize on performing more validation studies. Overall, this approach has a relativelyhigh accuracy of classifying surgical skill and it can be applied to different tasks in robotic-assisted surgery. [18]

B. Automating subtasks in robot-assisted laparoscopically surgery

1) Past and Current Minimally Invasive Surgery: Minimally Invasive Surgery (MIS) has become commonplace for an ever-growing number of operations. MIS is performed through small incisions or ports in the patient's body which has many advantages for the patient's compared to conventional "open" surgery. On the other side surgeons have to had cope with reduced dexterity and perception, which makes simple tasks such as knot-tying very time consuming. In conventional open surgeries tied suture knot take one second, compared to three minutes with laparoscopically tied knot. Until today robot-assisted MIS try to restore the feel of normal surgery by providing the surgeon with a more intuitive and ergonomic interface. The most common Robotic surgical systems today are DaVinci [21] and ZEUS [22]. Beside remarkable advances in robot-assisted surgery, some tasks like knot-tying are still inconvenient and time-consuming. Automating the subtask of knot-tying would be decrease the total surgery time. The past approaches in autonomous robotic knot-tying the controller uses a hard-wired policy. Which leads to the negative side effect that the controller repeats the same prescribed motion without the possibility of generalizing to unfamiliar instrument locations.

2) Using recurrent neural network for automating tie knots: One possible way to get a more robust control is to use deep learning to learn the control policy from examples of correct behaviour. One approach is on automating suture knot winding by training a recurrent neural network (RNN) introduced by Mayer et al. [23]. Unlike traditional neural network where we assume that all inputs and outputs are independent, RNNs have an internal state which means the output being depended on the previous computations [24]. For the knot-tying application previous states as the instrument position can be remembered for long periods of time in order to select future actions. Mayer gets the best results by using the hybrid supervised/evolutionary learning framework Evolino [25] and the Endoscopic Partial-Autonomous Robot (EndoPAR) system as shown in 'Fig. 10", which is an experimental robotic surgical platform developed by the Robotics and Embedded Systems research group at the Technical University of Munich. The speed-up in knot winding results in a total time of 25.8 sec for the entire knot. The preprogrammed controller needs 33.7 sec. The results show that RNNs can be useful assistance to speed up MIS procedures.

3) Outlook Automating subtasks in robot-assisted laparoscopically surgery with Deep learning: Supervised machine learning can be used to capture and generalize expertise without requiring the expansive controller design. In the future

Fig. 10. The EndoPAR system for tie knots in laparoscopically surgery. Source: [23]

the basic approach which is used for the knot-tying task could be also used for the thread tensioning performed by the assistant gripper, which is currently implemented by programmed controllers [23].

VI. Deep Learning using Convolutional Neural Networks Applied to Electromyography Data

During the last decade, Deep Learning (DP) and Convolutional Neural Networks (CNNs) have made it possible for many transformations and improvements on different fields where we apply Machine Leaning (MI) such as speech recognition and computer vision. Although their usage in the prosthetics field is uncommon, current research shows that CNNs can be beneficial in providing more control for prosthetic limbs. Here will be shown how simple CNNs can achieve above-average results when used to classify hand movements based on surface electromyography (sEGM).

1) Current status/situation (of prosthetic hands): Current methods used for controlling prosthetics hand based on surface electromyography and pattern recognition are promising. But even though the current commercially available prostatic hands are mechanically very advanced, they allow natural control for only a handful of movements and lack sufficient robustness and control speed. Pattern recognition algorithms are applied on the data captured by sEGM to classify the movement the amputee would have imagined doing, and an accuracy of more than 90 percentage can only be achieved on less than 10 classes.

2) sEGM Data used for Deep Learning: The classification of sEGM data using CNNs was done by [26]. The data are from the Ninapro database [27] and include electromyography data related to hand movements of 78 subjects (11 transradial amputees, 67 intact subjects) divided into three datasets. Dataset 1 included 27 intact subjects, Dataset 2 included 40 intact subjects, Dataset 3 included 11 transradial amputees. Several sensors were used to record hand kinematics, dynamics and correspondent muscular activity during the experiments. The setup configurations were the same for Dataset 2 and 3, but different for Dataset 1. The differences included type of electrodes, number of electrodes, the frequency of the raw sEGM signal. Data acquisitions were performed with two types of exercises. In one, the subjects imitated several repetitions of hand movements that were shown a screen, in the other the subjects repeated nine force patterns by pressing with one or more hand digits on a Finger-Force Linear Sensor. The acquisition protocol included several repetitions: 10 repetitions for Dataset 1, 6 repetitions for Dataset 2 and 3.

3) Classification using CNN compared to classical techniques: The convolutional neural network used by the research team for classification was a modified version of a simple and well known convolutional neural network [**?**]. The input data was divided in time frames of 150 ms, a choice comparable to what is commonly selected in the field. The CNN was chosen as such, in order to allow fast testing and assessment of different initial parameters and decisions such as in net architecture, optimization parameters and pre-processing. Afterwards, the data was classified using average traditional classifiers such as random Forests [28], SVM [29], and k-Nearest Neighbours [30].

TABLE I
Accuracy achieved at data classification

	CNN	Classical Methods
Dataset 1	66.59 ±6.40%	62.06 ±6.07%
Dataset 2	60.27 ±7.7%	60.28 ±6.51%
Dataset 3	38.09 ±14.29%	38.82 ±11.99%

As can be seen at Table 1, the accuracy achieved using an uncomplicated convolutional neural network is fairly comparable to the accuracy achieved by traditional classification methods. It is also worth noting that changes in different initial factors have a strong effect in the classification accuracy and the error rate of the convolutional neural network.

4) Conclusion: Convolutional neural networks with a simple architecture can achieve average and comparable results to traditional machine learning classification methods when analysing sEGM data. Based on literature [31] that showed that larger networks can achieve higher accuracy on different tasks, it is worth examining if larger CNNs can also improve the classification of sEGM data. An emerging technology, CNNs can provide useful in the prosthetics field, and therefore held to bring the idea of natural control of robotic hands one step closer to reality. (0)

VII. Learning control in robot-assisted rehabilitation of motor skills- a review [32]

With their economics and increased efficiency, robots nowadays are sought at as being highly attractive, thats why a lot of work and time is invested into incorporating these machines into the medical sphere. The loss of mobility as a consequence of a physical traumatic event, is a very serious subject in the medical world and the rehabilitation of the patients with the help of robots, as opposed to the current practice of using (just) a physiotherapist, is being researched intensively.

The present rehabilitation is a very time and energy consuming process, which also requires an experienced practitioner. To make matters worse, their number in the world is well below the wished lower bound. Here is where the robotic platforms come into play: they will be designed to augment the physical therapists output and to reach a lot more patients on the distance. The physical presence of the subjects in the practitioners office will no longer be needed; the robot, empowered with the knowledge and the best practices of physical therapy, but also with the help of technologies as machine learning, bio-mechanics etc. will conduct the rehabilitation process. Apart from the obvious good outcome, which is the rehabilitation of more patients, the economic advantages of such approach, even though the projected numbers are only on the early stages, promise great ROIs.

Even though that builds their basis, we are not interested in the robotic-tasks of the machines, but mostly in their ability to adopt to the very fragile motor skills of the injured individual (at the beginning of the rehabilitation) but also to their progress without the constant feedback of the physical therapist. Thats where a lot of the financial gain of this approach lies. The main idea is that a combo of subject tailored program and rich robotic algorithmic backup will enhance this branch of medicine. First lets define that the human species learn or relearn a task (even when the conditions are more disadvantageous, i.e. the loss of motor skills or their impediment from the traumatic injury) on a similar fashion. That is good news, because we can build a model from healthy subjects and map this model to the rehabilitation protocol of our impaired patients. In order to progress, we must first look where we stand in relation to technology-assisted rehabilitation. Because the subject is very wide and we dont want to loose the context by distributing our attention on many places, we will limit ourselves on the robotic systems that aid with upper limb impairments. We can categorize them as:

- Passive mechanisms equipped with an array of sensors, which measure parameters such as position, displacement or inertia. These mechanisms are also equipped with a virtual environment to guide the patients through different from the mechanism imposed challenges during the recovery phase.
- Robotic manipulanda, which measure hand motion and/or interaction forces, and thus regulates the movement of the subjects limb and the end-effector of the robotic machine.
- Robotic exosceletons, which assist the subjects limb at the joint region through exertion of adjusted force, that with the intention of providing more natural contact points. With the help of the mechanisms sensors (like joint angle, torque and position) the apparatus is capable of providing a more natural environment than the competition.

But what does the research say about the supremacy of robot-assisted rehabilitation versus the classical approach? In a research [33] the two approaches were compared on 73 patients and was concluded that the robot-assisted alternative was superior, but unfortunately by a very small margin. The same results were concluded from an other study [34]. But when looked closely, the studies had some flaws. In an attempt to keep the respective settings similar to each other, some factors that need to be considerably more emphasized for the robotic approach, like length and number of training sessions, were kept the same for both. An other major flow of the message derived from the conclusions, was that they were comparing robotic vs non-robotic, as if we were to completely eliminate the influence of a physical therapist. On the contrary, what we are trying to do is to enhance the work of the former through the introduction of these robotic systems. So the real question should be: In what extent has the robotic factor enhanced the quality of the rehabilitation process and changed the feedback loop from patient-physical therapist by introducing itself into a newly created patient-physical therapist-robot one. Such a question was partly answered on two studies [35] and [36]. It was concluded that robot-assisted systems offer, as expected, better measurements of the patients conditions and improvement during the rehabilitation phase. That provides a better to the patient needs and progress tailored recovery protocol, which in turn means faster progress and as a consequence more economic benefits.

But what are some of the differences between the two approaches, the conventional versus robot-assisted? In the current/conventional approach, the practitioner aids the patients by supporting their weight and exerting different force vectors into them, so they can relearn a movement through repetition. The experience of the physical therapist is a very important variable in this equation, because with the feedback of the patient as input, they can mold the way the rehabilitation will take place in the future. This to the patient adjusted and ever self-regulating force exertion can be done by the robotic system with just as, if not better, efficacy and in a much higher frequency than the human practitioner. The key here is to develop algorithms that auto regulate according to the feedback of the patient. This is the field of research where the majority of resources and attention is and will be invested. There are a few categories of control approaches and we list them as follows:

- Assistive forces: Aid the patients as little as possible in task completion.
- Resistive forces: Aid the patients in muscle hypertrophy (strengthening of the muscles) and endurance by guiding them through exercises that mimic a natural human movement.
- Corrective forces: Aid the patients in the pattern (re-)learning process by correcting incorrect movements.
- Disturbing forces: Aid the patients in the rehabilitation process by introduction of forces, that throw them gently of balance, so that a correction takes place from their part.

A robot is capable of doing all of these movements, but most importantly, it can do a combination of all four. Unfortunately, at the moment it doesn't exist a widely accepted control design methodology to model the rehabilitation of an individual. Physical therapists rely on their experience to execute the rehabilitation, and they use different ways of achieving the same end goal. There are variables like for example lactic acid accumulation in the muscles, which causes fatigue, range of motion of the patient, the trauma that caused the impediment and its severity etc. which have to be taken into account for such designs. The lack of a global definition and such different from each other scenarios cause unclarity. Thats why the researchers are using the ILC (Iterative Learning Control) paradigm, which in a nutshell is learning through repeated experimentation. A starting program as an initial point is designed and with the feedback acquired during the rehabilitation process, exercises and movement pattern to the initial array of the patterns will be added. A derivative of ILC is the windsurfer approach, which identifies and introduces new control design in a repetitive fashion throughout the whole process. So the idea is as the rehabilitation progresses, the same is done from the mobility.

As we can conclude from this (short) review of the current state of robot-assisted-rehabilitation is in its infancy stage right now. We are still trying to unify the field into a single universally accepted approach and a rehabilitation design model. This is indeed a very hard labor but with the giant steps of AI in the recent years, the worlds pattern recognition database will augment itself and thus help us in modeling individual rehabilitation programs for the patients of the (near) future.

References

[1] H. A. Pierson and M. S. Gashler, "Deep learning in robotics: a review of recent research," *Advanced Robotics*, vol. 31, no. 16, pp. 821–835, 2017.

[2] M. A. Nielsen, *Neural networks and deep learning*. Determination Press, 2015.

[3] J. Ker, L. Wang, J. Rao, and T. Lim, "Deep learning applications in medical image analysis," *IEEE Access Special Section*, 2017.

[4] S. Kannan, V. Gurusamy, and G. Nalini. (2014, October) Review on image segmentation techniques. [Online]. Available: https://www.researchgate.net/publication/273127438 review on image segmentation techniques

[5] K. Bhargavi and S. Jyothi, "A survey on threshold based segmentation technique," *Image Processing in International Journal of Innovative Research and Development*, vol. 3, November 2014.

[6] D. Kaur and Y. Kaur, "Various image segmentation techniques: A review," *International Journal of Computer Science and Mobile Computing*, vol. 3, pp. 809–814, May 2014.

[7] D. Bouget, M. Allen, D. Stoyanov, and P. Jannin, "Vison-based and marker-less surgical tool detection and tracking: a review of the literature," *Medical Image Analysis*, vol. 35, pp. 633–654, January 2017.

[8] Z. Zhao, S. Voros, Y. Weng, F. Chang, and R. Li, "Tracking-by-detection of surgical instruments in minimally invasive surgery via the convolutional neural network deep learning-based method," *Computer Assisted Surgery*, vol. 22, no. S1, pp. 26–35, 2017.

[9] L. Garcia-Peraza-Herrera, W. Li, C. Gruijthuijsen, A. Devreker, and G. Attilakos, "Real-time segmentation of non-rigig surgical tools based on deep learning and trackng," *Computer-Assisted and Robotic Endoscopy*, vol. 10170, pp. 84–95, February 2017.

[10] D. Pakhomov, V. Premachandran, M. allan, M. Azizian, and N. Navab. (2017, March) Deep residual learning for instrument segmentation in robotic surgery. [Online]. Available: https://arxiv.org/abs/1703.08580

[11] A. In, C. Direkolu, and M. ah, "Review of mri-based brain tumor image segmentation using deep learning methods," *Procedia Computer Science*, vol. 102, pp. 317 – 324, 2016, 12th International Conference on Application of Fuzzy Systems and Soft Computing, ICAFS 2016, 29-30 August 2016, Vienna, Austria. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S187705091632587X

[12] L. Chen, Y. Wu, A. M. DSouza, and A. Z. Abidin, "Mri tumor segmentation with densely connected 3d cnn," vol. 10574, 2018, pp. 10 574 – 10 574 – 8. [Online]. Available: https://doi.org/10.1117/12.2293394

[13] N. Porz, S. Bauer, A. Pica, P. Schucht, J. Beck, R. K. Verma, J. Slotboom, M. Reyes, and R. Wiest, "Multi-modal glioblastoma segmentation: Man versus machine," *PLOS ONE*, vol. 9, no. 5, pp. 1–9, 05 2014. [Online]. Available: https://doi.org/10.1371/journal.pone.0096873

[14] C. Roder, S. Bisdas, F. H. Ebner, J. Honegger, T. Naegele, U. Ernemann, and M. Tatagiba, "Maximizing the extent of resection and survival benefit of patients in glioblastoma surgery: High-field imri versus conventional and 5-ala-assisted surgery," *European Journal of Surgical Oncology*, vol. 40, no. 3, pp. 297–304, 2018/06/10 2014. [Online]. Available: http://dx.doi.org/10.1016/j.ejso.2013.11.022

[15] C. Doignon, F. Nageotte, and M. de Mathelin, "Segmentation and guidance of multiple rigid objects for intra-operative endoscopic vision," in *Dynamical Vision*, R. Vidal, A. Heyden, and Y. Ma, Eds. Berlin, Heidelberg: Springer Berlin Heidelberg, 2007, pp. 314–327.

[16] S. Petscharnig and K. Schöffmann, "Learning laparoscopic video shot classification for gynecological surgery," *Multimedia Tools and Applications*, vol. 77, no. 7, pp. 8061–8079, Apr 2018. [Online]. Available: https://doi.org/10.1007/s11042-017-4699-5

[17] G. H. Ballantyne and F. Moll, "The da vinci telerobotic surgical system: the virtual operative field and telepresence surgery," *Surgical Clinics*, vol. 83, no. 6, pp. 1293–1304, 2018/06/11 2003. [Online]. Available: http://dx.doi.org/10.1016/S0039-6109(03)00164-6

[18] Z. Wang and A. M. Fey, "Deep learning with convolutional neural network for objective skill evaluation in robot-assisted surgery," *arXiv preprint arXiv:1806.05796*, 2018.

[19] C. E. Reiley, H. C. Lin, D. D. Yuh, and G. D. Hager, "Review of methods for objective surgical skill evaluation," *Surgical endoscopy*, vol. 25, no. 2, pp. 356–366, 2011.

[20] Z. Wang and A. M. Fey, "Satr-dl: Improving surgical skill assessment and task recognition in robot-assisted surgery with deep neural networks," *arXiv preprint arXiv:1806.05798*, 2018.

[21] G. S. Guthart and J. K. Salisbury, "The intuitive/sup tm/ telesurgery system: overview and application," in *2000 IEEE international conference on robotics and automation*. IEEE, 2000, pp. 618–621.

[22] A. GARCIA-RUIZ, N. G. SMEDIRA, F. D. LOOP, J. F. HAHN, J. H. MILLER, C. P. STEINER, and M. GAGNER, "Robotic surgical instruments for dexterity enhancement in thoracoscopic coronary artery bypass graft," *Journal of Laparoendoscopic & Advanced Surgical Techniques*, vol. 7, no. 5, pp. 277–283, 1997.

[23] H. Mayer, F. Gomez, D. Wierstra, I. Nagy, A. Knoll, and J. Schmidhuber, "A system for robotic heart surgery that learns to tie knots using recurrent neural networks," in *2006 IEEE/RSJ International Conference on Intelligent Robots and Systems*. IEEE, 09/10/2006 - 15/10/2006, pp. 543–548.

[24] R. J. Williams and D. Zipser, "A learning algorithm for continually running fully recurrent neural networks," *Neural Computation*, vol. 1, no. 2, pp. 270–280, 1989.

[25] J. Schmidhuber, D. Wierstra, and F. Gomez, *Evolino: Hybrid Neuroevolution/Optimal Linear Search for Sequence Learning*, 2005.

[26] M. Atzori, M. Cognolato, and H. Müller, "Deep learning with convolutional neural networks applied to electromyography data: A resource for the classification of movements for prosthetic hands," *Frontiers in neurorobotics*, vol. 10, p. 9, 2016.

[27] M. Atzori, A. Gijsberts, C. Castellini, B. Caputo, A.-G. M. Hager, S. Elsig, G. Giatsidis, F. Bassetto, and H. Müller, "Electromyography data for non-invasive naturally-controlled robotic hand prostheses," *Scientific data*, vol. 1, p. 140053, 2014.

[28] L. Breiman, "Random forests," *Machine learning*, vol. 45, no. 1, pp. 5–32, 2001.

[29] N. Cristianini, J. Shawe-Taylor *et al.*, *An introduction to support vector machines and other kernel-based learning methods*. Cambridge university press, 2000.

[30] R. O. Duda, P. E. Hart, D. G. Stork *et al.*, "Pattern classification. 2nd," *Edition. New York*, vol. 55, 2001.
[31] Y. Bengio, I. J. Goodfellow, and A. Courville, "Deep learning," *Nature*, vol. 521, no. 7553, pp. 436–444, 2015.
[32] S.-H. Zhou, J. Fong, V. Crocher, Y. Tan, D. Oetomo, and I. Mareels, "Learning control in robot-assisted rehabilitation of motor skills–a review," *Journal of Control and Decision*, vol. 3, no. 1, pp. 19–43, 2016.
[33] V. Klamroth-Marganska, J. Blanco, K. Campen, A. Curt, V. Dietz, T. Ettlin, M. Felder, B. Fellinghauer, M. Guidali, A. Kollmar *et al.*, "Three-dimensional, task-specific robot therapy of the arm after stroke: a multicentre, parallel-group randomised trial," *The Lancet Neurology*, vol. 13, no. 2, pp. 159–166, 2014.
[34] A. C. Lo, P. D. Guarino, L. G. Richards, J. K. Haselkorn, G. F. Wittenberg, D. G. Federman, R. J. Ringer, T. H. Wagner, H. I. Krebs, B. T. Volpe *et al.*, "Robot-assisted therapy for long-term upper-limb impairment after stroke," *New England Journal of Medicine*, vol. 362, no. 19, pp. 1772–1783, 2010.
[35] V. S. Huang and J. W. Krakauer, "Robotic neurorehabilitation: a computational motor learning perspective," *Journal of neuroengineering and rehabilitation*, vol. 6, no. 1, p. 5, 2009.
[36] M. Casadio and V. Sanguineti, "Learning, retention, and slacking: a model of the dynamics of recovery in robot therapy," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 20, no. 3, pp. 286–296, 2012.