Supervised CNN strategies for optical image segmentation and classification in interventional medicine

Sara Moccia, Luca Romeo, Lucia Migliorelli, Emanuele Frontoni, Primo Zingaretti

Abstract The analysis of interventional images is a topic of high interest for the medical-image analysis community. Such an analysis may provide interventional-medicine professionals with both decision support and context awareness, with the final goal of improving patient safety. The aim of this chapter is to give an overview of some of the most recent approaches (up to 2018) in the field, with a focus on Convolutional Neural Networks (CNNs) for both segmentation and classification tasks. For each approach, summary tables are presented reporting the used dataset, involved anatomical region and achieved performance. Benefits and disadvantages of each approach are highlighted and discussed. Available datasets for algorithm training and testing and commonly used performance metrics are summarized to offer a source of information for researchers that are approaching the field of interventional-image analysis. The advancements in deep learning for medical-image analysis are involving more and more the interventional-medicine field. However, these advancements are undeniably slower than in other fields (e.g. preoperative-image analysis) and considerable work still needs to be done in order to provide clinicians with all possible support during interventional-medicine procedures.

Sara Moccia
Department of Information Engineering, Università Politecnica delle Marche, Ancona (Italy)
Department of Advanced Robotics, Istituto Italiano di Tecnologia, Genoa (Italy)
e-mail: s.moccia@univpm.it

Luca Romeo
Department of Information Engineering, Università Politecnica delle Marche, Ancona (Italy)
Department of Cognition, Motion and Neuroscience and Computational Statistics and Machine Learning, Istituto Italiano di Tecnologia, Genoa (Italy)
Department of Computational Statistics and Machine Learning, Istituto Italiano di Tecnologia, Genoa (Italy)
e-mail: l.romeo@univpm.it

Lucia Migliorelli, Emanuele Frontoni and Primo Zingaretti
Department of Information Engineering, Università Politecnica delle Marche, Ancona (Italy)
e-mail: l.migliorelli@pm.univpm.it, e.frontoni@univpm.it, zinga@dii.univpm.it

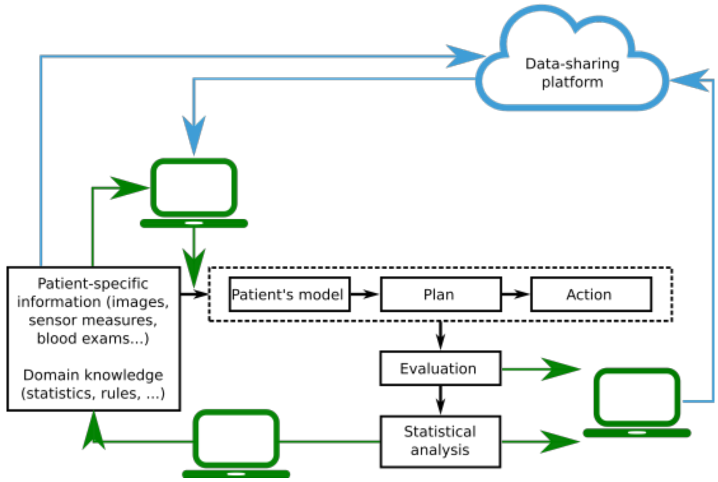

Fig. 1 Surgical data science integration in the interventional-medicine workflow allows objective decision-making and quantitative evaluation of the surgical outcomes.

1 Introduction to optical-image analysis in interventional medicine

Nowadays, the surgeon's decision process combines (i) pre-operative qualitative analysis of patient-specific anatomy and physiology, retrieved from imaging systems and sensors, and (ii) surgeon's prior knowledge about medical rules and statistics [1]. Such information is used to build an implicit patient's model and define a surgical plan. After-surgery, surgical outcomes are qualitatively evaluated and statistically analyzed to improve treatment effectiveness and eventually change treatment protocol (Fig. 1).

Advancements in intra-operative imaging systems and computer-based analysis allowed to acquire more and more information on patient's anatomy and physiology to eventually update the surgical plan directly in the operating room (OR). In fact, surgeons commonly exploit optical imaging when performing interventional-medicine procedures for obtaining both diagnostic support and context awareness in a non-invasive way [2]. New imaging devices that combine advanced sensors and increased computational power are constantly introduced in the OR, e.g., multispectral [3], narrow-band [4], and spectroscopy imaging [5]. Endoscopic cameras today allow to perform minimally invasive surgery (MIS) improving post-operative patient's prognosis and quality of life [1]. Robotic MIS is gradually emerging as a powerful solution to further improve treatment quality, and is already the state of the art in specific fields (e.g., urology) [6].

As a natural result of the massive introduction of imaging devices in the OR, an almost unlimited amount of electronic patient records are available [1]. These data can be processed in a quantitative way to further increase safety, effectiveness and

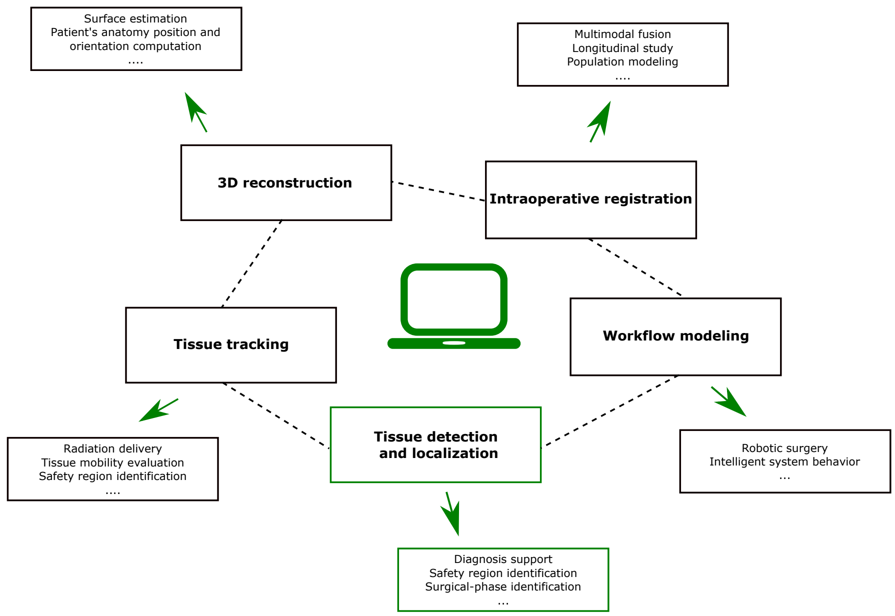

Fig. 2 Some of the major opportunities that surgical data science offers to interventional medicine. Blocks highlighted in green identify the main topics of this paper.

efficiency of surgical care [2]. Moreover, as observed in [7], the Internet-Of-Things revolution has the healthcare domain as one of the most promising field, with infinite opportunities arising from data sharing among hospitals, care-givers and patients. Indeed, data sharing can provide the surgeons with statistics from other patients shared among care centers and this information can integrate the patient-specific (local) data.

A primary goal of the medical image analysis community is to organize, analyze and model such huge amount of data to enhance the quality of interventional healthcare [2]. In this context, surgical data science (SDS) aims at supporting health specialists through a quantitative processing of intra-operative images to implement (Fig. 2): tissue tracking [8], 3D reconstruction [9], intra-operative registration [10], workflow modeling [11], detection and localization of anatomical structures [12] or/ and surgical instrumentation [13].

In addition to challenges related to intra- and inter-patient variability in biological tissues (especially in presence of pathologies), the processing of optical images acquired during interventional medicine presents further challenges, such as high sensor noise, varying illumination levels, organ movement, different pose of the acquisition sensor with respect to the tissues and presence of blood, smoke and surgical tools in the field of view.

To tackle the high variability of intra-operative optical images, SDS methods and principles heavily build on machine learning (ML) [2]. The medical domain-specific knowledge can be encoded in a ML-based model through a learning process based on the description of cases solved in the past. The model can:

- Offer decision support [11], e.g., by assisting the clinician when diagnosing new patients to improve the diagnostic speed, accuracy and/or reliability;
- Provide context awareness [14], e.g., for autonomous assistance and collaborative robots in MIS to improve safety, quality and efficiency of care.

More recently, deep learning (DL) approaches based on Convolutional Neural Networks (CNNs) for the analysis of interventional-medicine images drew the attention of the SDS community. Remarkable results were obtained in skin-cancer classification [15], polyp detection [16], retinal image analysis [17], and vessel segmentation [18], where large and labeled datasets are publicly available for DL model training. With respect to standard ML approaches to medical optical-image analysis, which require to extract high-level complex features (Sec. 1.2), CNNs tackle the classification and segmentation problems from a different point of view and represent the image as a nested hierarchy of simpler features that are automatically learned from the images during the training phase.

1.1 Aim of the survey

As the use of CNNs in the field of intra-operative optical image analysis is rapidly growing, the primary goal of this review is to provide an up-to-date source of information about its current state in the literature, with a specific focus on tissue classification and segmentation approaches for decision support and context awareness during interventional-medicine procedures.

Reviews in the field of DL for medical image analysis have been previously proposed, but mainly for applications related to anatomical images (such as computed-tomography or magnetic-resonance images) [19], while very few to describe the specific state of the art related to optical images acquired during interventional-medicine procedures. These latter ones only focus of specific anatomical regions without giving an integral vision of the challenges and advancements related to intra-operative tissue analysis. Examples include [20], that surveys methods for gastrointestinal-image analysis from a clinical point of view (more than from a methodological one). In [21] [22], algorithms for polyp and Barrett's esophagus detection are discussed, respectively, focusing on model-based and standard ML algorithms, leaving few space for DL strategies.

This survey may represent a salient resource for researchers in the field of SDS who wants to face up to the problem of intra-operative tissue analysis with DL. It analyzes almost fifty articles published from 2015, both from the methodological and application point of view. After a short introduction (Sec. 1.2) on last-decade methodologies, which mostly dealt with standard ML approaches, a short overview on CNNs is given (Sec. 1.3).

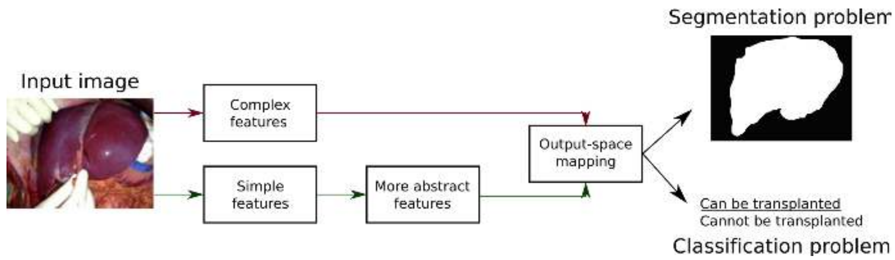

Fig. 3 Image segmentation and classification workflows for standard machine-learning (red arrows) and deep-learning (green arrows) approaches.

Considering the importance of having a proper and large training set to encode image and tissue variability when performing tissue classification and segmentation, a section to list and analyze the publicly available and labeled datasets is also included, along with a list of the most common metrics to evaluate algorithm performance in a fair and consistent way (Sec. 1.4).

CNN-based methodologies to image analysis are grouped in two categories: image segmentation (Sec. 2) and image classification (Sec. 3). As the majority of datasets built for interventional-medicine segmentation include also surgical-tool annotation, surgical-tool segmentation strategies are included in Sec. 2, too. In each category, articles are further split according to their clinical tasks. Finally, Sec. 4 concludes this paper summarizing the main findings and presenting open challenges and future research direction.

1.2 Previous approaches to tissue segmentation and classification

During the last decades, standard ML models for tissue classification typically applied (i) automated image analysis to extract a vector of quantitative, hand-designed, features to characterize the relevant image content and (ii) a pattern classifier to map the features to the output space to determine the category to which the extracted feature vector belongs, e.g., malignant/healthy tissue (Fig. 4).

The most exploited features were built from intensity, textural and derivative-based information [23]. Intensity-based features aimed at encoding information related to the prevalent intensity components in the image and were mainly based on intensity histogram, mean, variance and entropy. These features were commonly combined with textural features, which encoded tissue appearance, structure and arrangement [24]. Textural features included local binary patterns [25], gray-level co-occurence matrices [26] and histograms of oriented gradients [27]. Other popular features were obtained with filtering-based approaches, such as matched filtering and wavelet analysis,which have been widely used for polyp classification [28]. Similarly, derivative-based approaches built derivative-filters to extract image spatial derivatives, such as gradient and Laplacian, e.g., to highlight tissue edges [29].

As for pattern classifiers, several solutions were exploited. First attempts were based on probabilistic approaches (i.e., Naive Bayes) [30]. Similarly, perceptron-based algorithms have been widely used, e.g. for polyp detection in endoscopic images [31]. Tree-based algorithms and kernel based methods (i.e., support vector machine) were probably among the most widely used classifiers. These algorithms showed promising performance for tissue classification in several fields (e.g., abdominal-tissue segmentation and classification [24, 32, 26].

1.3 Background on convolutional neural networks (CNNs)

As in traditional neural networks, a CNN is a sequence of layers, where the convolutional one is the most peculiar. As pointed out in [33], convolution leverages three important ideas that can help improving classification and segmentation tasks with respect to traditional ML approaches (based on neural networks):

1. Sparse interactions. While for traditional networks every output unit interacts with every input unit, CNNs typically have sparse connections.
2. Parameter sharing. Rather than learning a separate set of features for every image location, only one set is learned, reasonably assuming that it is independent from the image location.
3. Equivariant representations. From the parameter-sharing property, the convolution equivariance to translation arises (i.e., if the input changes, the output changes in the same way).

Using convolutional layers results in fewer parameters to store and thus in reduced memory consumption, higher statistical efficiency and fewer operations to accomplish for output prediction.

In addition to convolution, CNNs commonly implement pooling between successive convolutional layers. With pooling, the output of the net at a certain location is replaced by a summary statistic of its nearby outputs (e.g., maximum value in case of max pooling). Implementing pooling is equivalent to perform downscaling, thus allowing noise smoothing and making the CNN invariant to small translations of the input.

Regarding image segmentation, today the most successful solutions exploit fully-convolutional neural networks (FCNNs), which allow a faster and more accurate segmentation. FCNNs were first presented in [34] and up to now several architectures, such as UNet [35], SegNet [36] and modified version of ResNet [37], showed remarkable segmentation performance.

For classification tasks, CNNs usually end with one or more fully-connected (dense) layers, i.e., layers where all the units have connection with the units of the previous layer (as in standard neural networks). The number of output units for the last layer coincides with the number of classes (e.g., two units for a binary classification problem such as healthy vs pathological tissue). From the first CNN model for image classification (i.e., LeNet5 [38]), today milestone architectures are

Alexnet [39], GoogleNet [40], VGG16 [41] and, more recently, fractal CNNs [42] and residual CNNs such as ResNet [43].

CNN based models were proposed for natural-image analysis, probably because of the availability of huge annotated datasets such as Imagenet ([1]). To take full advantage of the trained models (i.e., CNN weights) available online, a common strategy in interventional medicine imaging analysis is to implement fine tuning. Fine tuning consists in adapting the CNN weights learned with huge natural-image datasets by re-training the last CNN layers with the medical image dataset [15].

1.4 Available datasets and performance metrics

Considering the potentiality of learning algorithms to tackle the intra-operative image variability, collecting large quantity of annotated datasets for algorithm training became crucial. Indeed, several international organizations constantly work to collect and label, in a consistent manner, high-quality data recorded during interventional-medicine procedures. However, this positive trend still concerns only few anatomical regions (Table 1).

In parallel to the manual annotation of medical datasets, the SDS community is also studying how crowd- powered algorithm collaboration could be used to annotate large-scale medical images, as to moderate the surgeon involvement in the time-consuming annotation process [44].

Segmentation and classification performance is commonly evaluated with respect to the manual annotation performed by expert clinicians. To attenuate intra-subject variability when performing the manual annotation, a combination of annotation by multiple experts is usually employed [45]. When evaluating the algorithm performance with respect to manual annotation, a contingency table with true positive (TP), true negative (TN), false negative (FN) and false positive (FP) is commonly used. The positive and negative samples refer, in turn, to pixels within and outside the segmented region (segmentation task) or images belonging to diseased and healthy class (classification task) according to the manual annotation. Commonly exploited metrics that are computed from the contingency table are accuracy (Acc), sensitivity (Se), specificity (Sp) and precision (Pr):

$$Acc = \frac{TP + TN}{n} \tag{1}$$

$$Se = \frac{TP}{TP + FN} \tag{2}$$

$$Sp = \frac{TN}{TN + FP} \tag{3}$$

$$Pr = \frac{TP}{TP + FP} \tag{4}$$

[1] www.image-net.org/

Table 1 List of available datasets.

Name	Classification and segmentation task	Link
ISBI 2016, 2017	Celiac disease Gastric cancer Barrett's esophagus	https://aidasub-cleceliachy.grand-challenge.org/home/ https://aidasub-chromogastro.grand-challenge.org/ https://aidasub-clebarrett.grand-challenge.org/home/
KID	Gastrointestinal lesions	https://mdss.uth.gr/datasets/endoscopy/kid/
CVC colon DB	Colon polyps	http://mv.cvc.uab.es/projects/colon-qa/cvccolondb
EndoScene	Colon polyps	https://github.com/jbernoz/deeppolyp
MICCAI EndoVis	Colon polyps (ASU-Mayo) Vascular and inflammatory gastrointestinal lesions (GIANA) Outer edge of kidney (Kidney Boundary Detection), Barrett's esophagus (Early Barrett's cancer detection)	https://endovis.grand-challenge.org
Cervix dataset	Cervix type	https://www.kaggle.com/c/intel-mobileodt-cervical-cancer-screening/data
ISIC 2016, 2017, 2018	Dermoscopic images	https://www.isic-archive.com

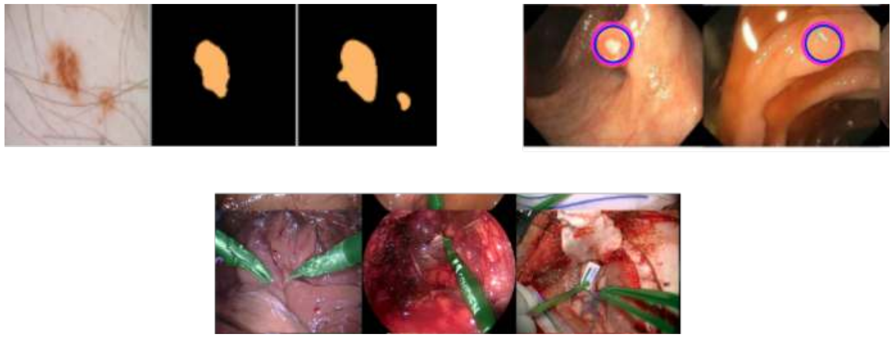

Fig. 4 Segmentation samples for skin, polyp and surgical instruments. Images adapted from [47, 57, 55].

being n the total number of pixels (segmentation task) or images (classification task).

The area under (AU) the Receiver Operating Characteristic (ROC) is also used as a metric (especially with skewed classes), where the ROC describes the performance of a binary classifier system as its discrimination threshold is varied.

When dealing with segmentation, further measures based on spatial overlapping can be used, too. The most used ones are the Dice Similarity Coefficient (DSC), also known as $F1\_score$, and the Jaccard coefficient (JC):

$$DSC = \frac{2TP}{FP + FN + 2TP} \tag{5}$$

$$JC = \frac{DSC}{2 - DSC} \tag{6}$$

2 Optical-image segmentation

This section will survey approaches for the segmentation of images acquired during interventional-medicine procedures. For each segmentation approach, Table 2 lists the relative anatomical region, image dataset, segmentation task and performance metrics. Figure 4 shows visual samples for skin, polyp and surgical-tool analysis.

Skin lesions

Following the first CNN-based approach to pathological skin-image analysis, mainly dealing with classification tasks [15], several methods for lesion segmentation have been proposed. In [47], an encoder-decoder network is proposed to melanoma segmentation. The network is based on U-Net and includes skip connections, as in ResNets, and dilated convolution [58]. ResNet is also used in [46]. A similar ap-

Table 2 Summary table for image-segmentation approaches. *Acc*: accuracy, *Se*: sensitivity, *Sp*: specificity, *Pr*: precision, *DSC*: Dice similarity coefficient, *JC*: Jaccard coefficient.

Article	Anatomical region	Dataset	Task	Performance metrics					
				Acc	*Se*	*Sp*	*Pr*	*DSC*	*JC*
Bi et al., 2017 [46]	Skin	ISIC 2017	Lesions						0.794
Sarker et al., 2018 [47]	Skin	ISBI 2016	Lesions	0.984	0.945	0.992		0.955	0.913
	Skin	ISBI 2017	Lesions	0.936	0.816	0.983		0.878	0.782
Mirikharaji et al., 2018 [48]	Skin	ISBI 2017	Lesions	0.938	0.855	0.973		0.857	0.773
Ghosh et al., 2018 [49]	Colon	KID	Lesions	0.944					
Wickstrom et al., 2018 [50]	Colon	EndoScene	Polyps	0.949					
Vazquez et al.,2016 [51]	Colon	EndoScene	Polyps	0.930					
Brandao et al.,2018 [52]	Colon	MICCAI EndoVis and CVC colon DB	Polyps						
Laina et al., 2017 [53]	Gastrointestinal tract	MICCAI EndoVis	Surgical tool	0.926	0.862	0.99		0.889	
Attia et al., 2017 [54]	Gastrointestinal tract	MICCAI EndoVis	Surgical tool	0.933					0.827
Garcia et al., 2015 [55]	Gastrointestinal tract	MICCAI EndoVis (non real time)	Surgical tool	0.837	0.722	0.952			
	Gastrointestinal tract	MICCAI EndoVis (real time)	Surgical tool	0.883	0.878	0.887			
Milletari et al. 2018 [56]	Gastrointestinal tract	MICCAI EndoVis	Surgical tool	0.978	0.888	0.988		0.895	

proach is proposed in [48], with the main innovation of including shape priors in the loss function used to train the FCNN. This yields to faster convergence and more accurate segmentation results. U-Net is also exploited in [59], where a nested architecture is proposed by optimizing a loss function that allows handling partial image labeling in confocal microscopy skin images.

Gastrointestinal lesions al polyps

A benchmark analysis for FCNN-based polyp segmentation is proposed in [51], using one of the first FCNN model in the literature [34]. In [50, 49], a modified version of SegNet is proposed for pixel-wise polyp and bleeding segmentation in wireless-endoscopy images, respectively. Polyp detection is achieved with SegNet in [49], too. In [50] a similar approach is investigated for polyp detection, with further segmentation-uncertainty estimation via Monte Carlo dropout and model interpretability analysis by highlighting descriptive regions in the input images with guided backpropagation [60].

Two parallel custom-built CNNs (for edge detection and lesion classification) are described in [61] to allow Hookworm disease detection in wireless endoscopic images. In [57], temporal information is included in the polyp detection process by building a 3D CNN. Experimental results show an improvement in the detection performance with respect to approaches based on single-frame processing.

Depth information is exploited in [52] as an additional input channel to FCNN architectures based on VGG16 and Resnet to the RGB information, experimentally demonstrating improved performance. Growing interest in the field is also reserved to automatic depth prediction with CNNs for 3D colon-shape reconstruction [62, 63, 64].

Surgical tools for gastrointestinal surgery

One of the first real-time FCNN-based approaches to the segmentation of non-rigid surgical tools was proposed in [55], where SegNet was adapted and fine-tuned to segment surgical tool in endoscopic images. A similar approach is proposed in [53], where the FCNN encoder is inspired by ResNet, and the decoder one has two branches for generating both the instrument segmentation mask and its articulated 2D pose.

In [86], a U-Net based architecture to surgical tool segmentation is proposed. The FCNN is modified to allow multiple instrument segmentation. The FCNN is in series with a second regressor network to regress the instrument pose.

Recurrent networks are used in [54, 56], where an encoder-decoder FCNN inspired to U-Net is combined with Long Short Term Memory (LSTM) to provide instrument segmentation in endoscopic images while encoding temporal dependencies. This methodology results in higher accuracy than approaches based on non-

Table 3 Summary table for image-classification approaches. WCE: wireless capsule endoscopy, *Acc*: accuracy, *Se*: sensitivity, *Sp*: specificity, *Pr*: precision, *AUC*: area under the receiver operating characteristic, *DSC*: Dice similarity coefficient.

Article	Anatomical region	Dataset	Task	n of classes	Performance metrics					
					Acc	*Se*	*Sp*	*Pr*	*AUC*	*DSC*
Esteva et al., 2017 [15]	Skin	129450 images	Lesions	2					0.953	
Bi et al., 2017 [46]	Skin	ISIC 2017	Lesions	2					0.855	
	Skin	ISIC 2017	Lesions	2+3 (ensemble)					0.976	
Zhang et al., 2018 [65]	Skin	ISIC 2016	Lesions	2	0.858			0.664	0.818	
Lopez et al., 2017 [66]	Skin	ISIC	Lesions	2		0.787		0.797		
Navarro et al., 2018 [67]	Skin	1300 images	Lesions	10					0,9695	
Pal et al., 2016 [68]	Skin	707 images	Lesions	3	0.589					
Mendel et al., 2017 [69]	Gastrointestinal tract	MICCAI EndoVis 2015	Lesions	2		0.940	0.880			
Georgakopoulos et al., 2016 [70]	Gastrointestinal tract	KID	Lesions	2	0.902	0.926	0.889			
Hong et al., 2017 [71]	Gastrointestinal tract	ISBI 2016	Lesions	3	0.808					
Ribeiro et al., 2016 [72]	Gastrointestinal tract	818 images	Polyps	2	0.936					
Yuan et al., 2018 [73]	Gastrointestinal tract	3000 WCE images	Polyps	2	0.956	0.950		0.963		0.956
Aoki et al.,2018 [74]	Gastrointestinal tract	15800 images	Lesions	2	0.908	0.882	0.909		0.958	
Fan et al.,2018 [75]	Gastrointestinal tract	WCE images	Lesions	2	0.953		0.954		0.909	
Sekuboyina et al.,2017 [76]	Gastrointestinal tract	137 images	Lesions	8					0.8	
Segui et al.,2016 [77]	Gastrointestinal tract	120K WCE images	Lesions	6	0.96					
Vasilakakis et al., 2018 [78]	Gastrointestinal tract	KID	Lesions	2	0.9					
Itoh et al., 2018 [79]	Gastrointestinal tract	170 images	Other applications (helicobapter pilori)	2		0.867	0.867		0.956	
Yu et al., 2015 [80]	Gastrointestinal tract	1mln real WCE images	Other applications (digestive organs)	3	0.973					
Chen et al., 2017 [81]	Gastrointestinal tract	630k images	Other applications (digestive organs)	3	0.897	0.943	0.900			
Aubreville et al., 2017 [82]	Oral cavity	7894 images	Other applications (cancerous lesions)	2	0.883	0.866	0.900		0.960	
Xu et al., 2016 [83]	Cervix	690 images	Other applications (cervix dysplasia)	2	0.889	0.878	0.900		0.940	
Zou et al., 2015 [84]	Gastrointestinal tract	1 mln real WCE images	Other applications (organ classification)	3	0.955					
Zhou et al., 2017 [85]	Gastrointestinal tract	8800 images	Other applications (celiac disease)	2		1	1			

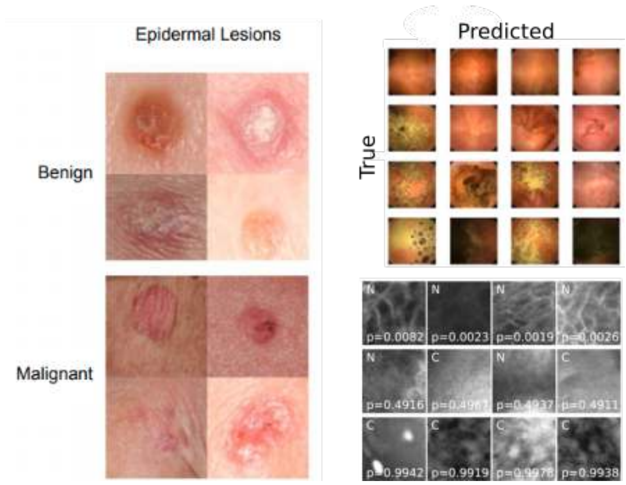

Fig. 5 Classification samples for skin lesions, polyp and oral-cavity cancer. Images adapted from [15, 77, 88].

recurrent networks. With the same aim, CNNs with 3D kernels have been proposed in [87] for instrument pose estimation.s

3 Optical-image classification

This section will survey approaches for the classification of images acquired during interventional-medicine procedures. For each segmentation approach, Table 3 lists the relative anatomical region, image dataset, classification task and performance metrics. Figure 5 shows visual samples for skin, gastrointestinal and oral-cavity lesion classification.

First approaches to CNN-based tissue classification exploited CNN simply to extract learned features, which then will be used for tissue classification with standard ML-approaches introduced in Sec. 1.2 [19]. This was mainly related to the small numerosity of image datasets. When larger datasets started to become publicly available, more advanced approaches were investigated, which we will survey hereafter. Accordingly, CNN-based approaches started to be exploited in order to (i) learn discriminative nonlinear features and (ii) classify the optical-images according to such features.

Skin lesions

The work presented in [15] is one of the first approaches to skin-lesion segmentation with CNN, where Google Inception v3 is fine-tuned to detect tumoral skin lesions. A similar approach, which uses VGG16 as classification network, is presented in [66], while ResNet is fine-tuned to classify skin lesions in [46]. A three-branch CNN is proposed in [68] for coarse classification of psoriatic-plaque macro classes. After a common VGG16-like architecture, each branch is deputy to a finer classification of plaque grade. Webly supervised learning is investigated in [67] to deal with data imbalance. An innovative approach to synergic DL has been recently proposed in [65], overcoming state of the art approaches.

Gastrointestinal lesions and polyps

Ulcer and bleeding in wireless endoscopic images are classified in [74, 76] using a sixteen- and ten-layer CNN, respectively. A similar approach is exploited in [77], where the CNN is fed with both endoscopic frames and image priors (Hessian and Laplacian) to improve the classification performance. For the same task, AlexNet is used in [75]. Interesting approaches to weakly-supervised CNN for detection of inflammatory gastrointestinal lesions are proposed in [70, 78], to overcome the problem of limited number of annotated images.

A simple CNN with six stages is used in [71] to classify Barrett's esophagus and neoplasia in endomicroscopy images. In [69] a further advancement is done and Barretts esophagus frames are classified by fine-tuning ResNet.

One of the first approaches to polyp classification using an end-to-end trained CNN is proposed in [72], where transfer learning is applied to several CNN models, such as VGG16 and Alexnet, outperforming conventional ML methods. An innovative approach to polyp classification is proposed in [73] where a 10-stages CNN architecture that consists of alternated convolutional and dense layers is built and regularized to be rotation-invariant.

Other applications

CNNs inspired by AlexNet are used in [81, 84] to identify digestive organs. A seven-stage CNN is used in [80] to automatically extract image features that are then classified with extreme ML. Such approach experimentally shows better performance than using a fully-connecting layers, probably due to the small depth of the CNN.

GoogleNet is used in [79], to classify Helicobacter pylori infection in upper gastrointestinal endoscopy images. Fine-tuning is implemented to transfer the recognition capabilities of the GoogleNet to the endoscopic images. A similar approach is used in [85] for celiac disease classification by video endoscopy.

A seven-stages CNN is tested in [82] to classify cancerous tissue in laserendomicroscopy images of the oral cavity, showing improved performance with respect to standard ML-based approaches in the field.

A multimodal CNN-based approach to cervical dysplasia classification is proposed in [83]: this combines both automatic feature extraction with CNNs and data from clinical records.

4 Discussion

The efforts in the field of DL applied to optical-image analysis are promising and encouraging, however, several methodological and technical challenges are still open, hampering the translation of these developed methodologies into the actual clinical practice.

From the methodological point of view, it emerged that a comparison of the proposed methodologies is not trivial. There is not a consensus yet on the exploited datasets and the reported performance metrics, which are not consistent among different research articles (see Table 2 and Table 3). Moreover, despite the efforts invested in the analysis of interventional-medicine images, the number of research articles in this field is still lower than that relate to anatomical-image analysis [19].

Regarding the technical challenges, there are several aspects that can be tackled to potentially achieve the goal of robust and reliable tissue classification.

The first aspect deals with hardware design. Indeed, the imaging field is constantly evolving thanks to new optical imaging technologies, such as narrow-band imaging [89] and multispectral imaging [90]. These technologies potentially allow high-quality optical imaging (e.g., in terms of image noise and tissue-background contrast) and have already found interesting applications in the remote-sensing field [91]. However, the use of these technologies is still underrepresented in the medical field with few examples (e.g. [32, 92]).

A second aspect is related to the identification of images to be processed. High noise level in the image, camera movements, tissue deformation and illumination drop lower image quality and make the classification challenging also for the human eye. Similarly, classification algorithms are prone to error when processing uninformative and noisy frames. Solutions have been proposed in the literature, nonetheless they are still limited to few anatomical regions and have to be further investigated [93, 88].

A third point concerns the estimation of the level of classification confidence while increasing the model interpretability, with a view to improve generalization performance. In particular, it has been reported that allowing a system to produce “unknown” results can potentially reduce the number of incorrectly classified cases [94]. In this context, advancements in DL aim to discover patterns sometimes unsighted by physicians [95] while estimating the posterior probability of the prediction. Interestingly, understanding why and how the outcome prediction is made may also help the physician to discover salient predictors involved in the diagnostic process (pattern localization) [96]. However, the introduction of confidence estimation in the medical imaging field has been only marginally explored [32]. DL-model interpretability is still an open research topic and recent approaches aim to increase it by (i) employing sparse CNN models with different loss or penalty functions [97] and (ii) exploiting visual-attention models to predict human eye fixations on images [98].

More generally, as SDS/DL strongly rely upon labeled data, the last aspect is related to the availability of labeled datasets. Indeed, the larger the training dataset, the bigger the chances the classification algorithm will be accurate in classifying unseen data. While the development of tissue-classification algorithms is strongly advancing in some specific fields (e.g., vascular district [18], and gastrointestinal tract [22]), there are other fields that are incredibly underrepresented in the literature. The most probable reason for this is indeed the lack of large and available labeled databases for algorithm training.

Despite international organizations are active in collecting high-quality annotated datasets, several anatomical districts are still underrepresented, thus limiting the applicability of supervised CNN-based approaches. However, only a fraction of patient-related data is digitized and stored in a structured and standardized way, and data quality assessment is rarely performed [2]. This is probably the main reason why SDS only recently emerged as an active field of research. While shared databases are available for other research fields for advancing research (e.g., the ImageNet dataset, www.image-net.org/), annotated datasets for the SDS community are still limited in number. This can be attributed to regulatory and sociological factors (e.g. data protection and privacy issues) [99]. A second factor deals with medical data annotation, which is typically an expensive process in terms of resources and time [100]. In the last years, several efforts have been made by the SDS community to support research-data sharing and develop crowd- powered algorithms for large-medical-dataset tagging [101, 102]. With a focus on imaging data, data sharing is especially supported by international organizations, such as the MICCAI society, the IEEE Signal Processing Society and the IEEE Engineering in Medicine and Biology Society, which yearly organize *Grand Challenges*[2] and release annotated dataset for algorithm testing (despite focusing mostly on anatomical imaging). However, when analyzing the description of datasets in Table 1, it emerged that information related to the number of patients / surgeries / healthcare centers involved in the dataset creation may be missing. This information could provide useful hints to be exploited by researcher when developing and testing DL algorithms (e.g. in terms of robustness to intra- and inter-patient variability) [103]. It is also worth noticing that the dataset numerosity (both in terms of images and patients) heavily varies from dataset to dataset (for each different clinical task). For example, in the MICCAI EndoVis dataset for small-bowel lesion localization, approximately 3600 images are given, while for early Barrett's cancer detection only 100 images are available.

Researchers are currently trying to overcome the DL shortcoming of requiring huge annotation datasets with unsupervised approaches where the problem of high dimensionality of the random variables to be modeled arise [33].

Even in presence of a sufficiently large labeled dataset, CNN training may not be trivial if the training labels are sparse, unbalanced or if there is not a consensus among health-operator annotations (e.g., in the definition of tumor margins). Specific weakly-supervised learning techniques, as multiple instance learning [104], may be used to address the problem of both temporal and spatial sparse labeling [105]. Solutions to face data unbalanced should be applied both at data and algorithm level [106], especially when training data are strongly unbalanced (i.e.,

[2] https://grand-challenge.org/all_challenges/

number of positive cases $\ll$ number of negative cases). In order to improve the annotation procedure, ranking algorithms [107, 108] can be used to sort the different responses of health-operator annotators, while evaluating the confidence level of the reported label.

In conclusion, to allow the actual integration of quantitative intra-operative image analysis into the actual clinical practice [109], the goal is developing adequate data-analysis technology to provide surgeons with quantitative support and effectively translate the technology into patient care workflow. SDS plays an important role in moving from (surgeon-specific) subjective to (computer-assisted) objective decision-making and from qualitative to quantitative assessment of surgical outcomes [2]. The integration of computer-aids will facilitate the surgeon's decision process and risk assessment, offering situation awareness, improved ergonomics and reduced cognitive workload.

References

1. Taylor, R.H., Menciassi, A., Fichtinger, G., Fiorini, P., Dario, P.: Medical robotics and computer-integrated surgery. In: Springer Handbook of Robotics, pp. 1657–1684. Springer (2016)
2. Maier-Hein, L., Vedula, S.S., Speidel, S., Navab, N., Kikinis, R., Park, A., Eisenmann, M., Feussner, H., Forestier, G., Giannarou, S., et al.: Surgical data science for next-generation interventions. Nature Biomedical Engineering **1**(9), 691 (2017)
3. Stewart, J.W., Akselrod, G.M., Smith, D.R., Mikkelsen, M.H.: Toward multispectral imaging with colloidal metasurface pixels. Advanced Materials **29**(6) (2017)
4. Machida, H., Sano, Y., Hamamoto, Y., Muto, M., Kozu, T., Tajiri, H., Yoshida, S.: Narrow-band imaging in the diagnosis of colorectal mucosal lesions: a pilot study. Endoscopy **36**(12), 1094–1098 (2004)
5. Emsley, J.W., Lindon, J.C.: NMR spectroscopy using liquid crystal solvents. Elsevier (2018)
6. Abbou, C.C., Hoznek, A., Salomon, L., Olsson, L.E., Lobontiu, A., Saint, F., Cicco, A., Antiphon, P., Chopin, D.: Laparoscopic radical prostatectomy with a remote controlled robot. The Journal of Urology **197**(2), S210–S212 (2017)
7. Balmer, J.M., Yen, D.A.: The internet of total corporate communications, quaternary corporate communications and the corporate marketing internet revolution. Journal of Marketing Management **33**(1-2), 131–144 (2017)
8. Stoyanov, D.: Surgical vision. Annals of Biomedical Engineering **40**(2), 332–345 (2012)
9. Maier-Hein, L., Mountney, P., Bartoli, A., Elhawary, H., Elson, D., Groch, A., Kolb, A., Rodrigues, M., Sorger, J., Speidel, S., et al.: Optical techniques for 3D surface reconstruction in computer-assisted laparoscopic surgery. Medical Image Analysis **17**(8), 974–996 (2013)
10. Sotiras, A., Davatzikos, C., Paragios, N.: Deformable medical image registration: A survey. IEEE Transactions on Medical Imaging **32**(7), 1153–1190 (2013)
11. März, K., Hafezi, M., Weller, T., Saffari, A., Nolden, M., Fard, N., Majlesara, A., Zelzer, S., Maleshkova, M., Volovyk, M., et al.: Toward knowledge-based liver surgery: Holistic information processing for surgical decision support. International Journal of Computer Assisted Radiology and Surgery **10**(6), 749–759 (2015)
12. Moccia, S., Foti, S., Routray, A., Prudente, F., Perin, A., Sekula, R.F., Mattos, L.S., Balzer, J.R., Fellows-Mayle, W., De Momi, E., et al.: Toward improving safety in neurosurgery with an active handheld instrument. Annals of Biomedical Engineering **46**(10), 1450–1464 (2018)
13. Nosrati, M.S., Peyrat, J.M., Abinahed, J., Al-Alao, O., Al-Ansari, A., Abugharbieh, R., Hamarneh, G.: Efficient multi-organ segmentation in multi-view endoscopic videos us-

ing pre-operative priors. In: International Conference on Medical Image Computing and Computer-Assisted Intervention, pp. 324–331. Springer (2014)
14. Katić, D., Schuck, J., Wekerle, A.L., Kenngott, H., Müller-Stich, B.P., Dillmann, R., Speidel, S.: Bridging the gap between formal and experience-based knowledge for context-aware laparoscopy. International Journal of Computer Assisted Radiology and Surgery **11**(6), 881–888 (2016)
15. Esteva, A., Kuprel, B., Novoa, R.A., Ko, J., Swetter, S.M., Blau, H.M., Thrun, S.: Dermatologist-level classification of skin cancer with deep neural networks. Nature **542**(7639), 115 (2017)
16. Bernal, J., Tajkbaksh, N., Sánchez, F.J., Matuszewski, B.J., Chen, H., Yu, L., Angermann, Q., Romain, O., Rustad, B., Balasingham, I., et al.: Comparative validation of polyp detection methods in video colonoscopy: Results from the MICCAI 2015 endoscopic vision challenge. IEEE Transactions on Medical Imaging **36**(6), 1231–1249 (2017)
17. Poplin, R., Varadarajan, A.V., Blumer, K., Liu, Y., McConnell, M.V., Corrado, G.S., Peng, L., Webster, D.R.: Prediction of cardiovascular risk factors from retinal fundus photographs via deep learning. Nature Biomedical Engineering p. 1 (2018)
18. Moccia, S., De Momi, E., El Hadji, S., Mattos, L.S.: Blood vessel segmentation algorithmsreview of methods, datasets and evaluation metrics. Computer Methods and Programs in Biomedicine **158**, 71–91 (2018)
19. Litjens, G., Kooi, T., Bejnordi, B.E., Setio, A.A.A., Ciompi, F., Ghafoorian, M., Van Der Laak, J.A., Van Ginneken, B., Sánchez, C.I.: A survey on deep learning in medical image analysis. Medical Image Analysis **42**, 60–88 (2017)
20. Patel, V., Armstrong, D., Ganguli, M., Roopra, S., Kantipudi, N., Albashir, S., Kamath, M.V.: Deep learning in gastrointestinal endoscopy. Critical Reviews in Biomedical Engineering **44**(6) (2016)
21. Prasath, V.B.S.: Polyp detection and segmentation from video capsule endoscopy: A review. Journal of Imaging **3**(1) (2017)
22. de Souza, L.A., Palm, C., Mendel, R., Hook, C., Ebigbo, A., Probst, A., Messmann, H., Weber, S., Papa, J.P.: A survey on Barrett's esophagus analysis using machine learning. Computers in Biology and Medicine (In press)
23. Zhang, J., Xia, Y., Xie, Y., Fulham, M., Feng, D.D.: Classification of medical images in the biomedical literature by jointly using deep and handcrafted visual features. IEEE Journal of Biomedical and Health Informatics **22**(5), 1521–1530 (2018)
24. Zhang, Y., Wirkert, S.J., Iszatt, J., Kenngott, H., Wagner, M., Mayer, B., Stock, C., Clancy, N.T., Elson, D.S., Maier-Hein, L.: Tissue classification for laparoscopic image understanding based on multispectral texture analysis. Journal of Medical Imaging **4**(1), 015,001–015,001 (2017)
25. Misawa, M., Kudo, S.e., Mori, Y., Takeda, K., Maeda, Y., Kataoka, S., Nakamura, H., Kudo, T., Wakamura, K., Hayashi, T., et al.: Accuracy of computer-aided diagnosis based on narrow-band imaging endocytoscopy for diagnosing colorectal lesions: comparison with experts. International Journal of Computer Assisted Radiology and Surgery pp. 1–10 (2017)
26. Moccia, S., De Momi, E., Guarnaschelli, M., Savazzi, M., Laborai, A., Guastini, L., Peretti, G., Mattos, L.S.: Confident texture-based laryngeal tissue classification for early stage diagnosis support. Journal of Medical Imaging **4**(3), 034,502 (2017)
27. Freeman, W.T., Roth, M.: Orientation histograms for hand gesture recognition. In: International Workshop on Automatic Face and Gesture Recognition, vol. 12, pp. 296–301 (1995)
28. Magoulas, G.D.: Neuronal networks and textural descriptors for automated tissue classification in endoscopy. Oncology Reports **15**(4), 997–1000 (2006)
29. Kumar, S., Saxena, R., Singh, K.: Fractional fourier transform and fractional-order calculus-based image edge detection. Circuits, Systems, and Signal Processing **36**(4), 1493–1513 (2017)
30. Mukherjee, R., Manohar, D.D., Das, D.K., Achar, A., Mitra, A., Chakraborty, C.: Automated tissue classification framework for reproducible chronic wound assessment. BioMed Research International **2014** (2014)

31. Karargyris, A., Bourbakis, N.: Wireless capsule endoscopy and endoscopic imaging: A survey on various methodologies presented. IEEE Engineering in Medicine and Biology Magazine **29**(1), 72–83 (2010)
32. Moccia, S., Wirkert, S.J., Kenngott, H., Vemuri, A.S., Apitz, M., Mayer, B., De Momi, E., Mattos, L.S., Maier-Hein, L.: Uncertainty-aware organ classification for surgical data science applications in laparoscopy. IEEE Transactions on Biomedical Engineering **158**(65), 2649 – 2659 (2018)
33. Goodfellow, I., Bengio, Y., Courville, A., Bengio, Y.: Deep learning, vol. 1. MIT Press Cambridge (2016)
34. Long, J., Shelhamer, E., Darrell, T.: Fully convolutional networks for semantic segmentation. In: IEEE Conference on Computer Vision and Pattern Recognition, pp. 3431–3440 (2015)
35. Ronneberger, O., Fischer, P., Brox, T.: U-net: Convolutional networks for biomedical image segmentation. In: International Conference on Medical Image Computing and Computer-Assisted Intervention, pp. 234–241. Springer (2015)
36. Badrinarayanan, V., Kendall, A., Cipolla, R.: Segnet: A deep convolutional encoder-decoder architecture for image segmentation. arXiv preprint arXiv:1511.00561 (2015)
37. Drozdzal, M., Vorontsov, E., Chartrand, G., Kadoury, S., Pal, C.: The importance of skip connections in biomedical image segmentation. In: Deep Learning and Data Labeling for Medical Applications, pp. 179–187. Springer (2016)
38. LeCun, Y., Bottou, L., Bengio, Y., Haffner, P.: Gradient-based learning applied to document recognition. Proceedings of the IEEE **86**(11), 2278–2324 (1998)
39. Krizhevsky, A., Sutskever, I., Hinton, G.E.: Imagenet classification with deep convolutional neural networks. In: Advances in Neural Information Processing Systems, pp. 1097–1105 (2012)
40. Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., Anguelov, D., Erhan, D., Vanhoucke, V., Rabinovich, A.: Going deeper with convolutions. In: IEEE Conference on Computer Vision and Pattern Recognition, pp. 1–9 (2015)
41. Simonyan, K., Zisserman, A.: Very deep convolutional networks for large-scale image recognition. arXiv preprint arXiv:1409.1556 (2014)
42. Larsson, G., Maire, M., Shakhnarovich, G.: Fractalnet: Ultra-deep neural networks without residuals. arXiv preprint arXiv:1605.07648 (2016)
43. He, K., Zhang, X., Ren, S., Sun, J.: Deep residual learning for image recognition. In: IEEE Conference on Computer Vision and Pattern Recognition, pp. 770–778 (2016)
44. Heim, E., Roß, T., Seitel, A., März, K., Stieltjes, B., Eisenmann, M., Lebert, J., Metzger, J., Sommer, G., Sauter, A.W., et al.: Large-scale medical image annotation with crowd-powered algorithms. Journal of Medical Imaging **5**(3), 034,002 (2018)
45. Warfield, S.K., Zou, K.H., Wells, W.M.: Simultaneous truth and performance level estimation (STAPLE): an algorithm for the validation of image segmentation. Transactions on Medical Imaging **23**(7), 903–921 (2004)
46. Bi, L., Kim, J., Ahn, E., Feng, D.: Automatic skin lesion analysis using large-scale dermoscopy images and deep residual networks. arXiv preprint arXiv:1703.04197 (2017)
47. Sarker, M., Kamal, M., Rashwan, H.A., Banu, S.F., Saleh, A., Singh, V.K., Chowdhury, F.U., Abdulwahab, S., Romani, S., Radeva, P., et al.: SLSDeep: Skin lesion segmentation based on dilated residual and pyramid pooling networks. In: International Conference on Medical Image Computing and Computer-Assisted Intervention, pp. 21–29. Springer (2018)
48. Mirikharaji, Z., Hamarneh, G.: Star shape prior in fully convolutional networks for skin lesion segmentation. In: A.F. Frangi, J.A. Schnabel, C. Davatzikos, C. Alberola-López, G. Fichtinger (eds.) Medical Image Computing and Computer Assisted Intervention, pp. 737–745. Springer International Publishing, Cham (2018)
49. Ghosh, T., Li, L., Chakareski, J.: Effective deep learning for semantic segmentation based bleeding zone detection in capsule endoscopy images. In: IEEE International Conference on Image Processing, pp. 3034–3038. IEEE (2018)
50. Wickstrøm, K., Kampffmeyer, M., Jenssen, R.: Uncertainty modeling and interpretability in convolutional neural networks for polyp segmentation. In: International Workshop on Machine Learning for Signal Processing, pp. 1–6. IEEE (2018)

51. Vázquez, D., Bernal, J., Sánchez, F.J., Fernández-Esparrach, G., López, A.M., Romero, A., Drozdzal, M., Courville, A.: A benchmark for endoluminal scene segmentation of colonoscopy images. Journal of Healthcare Engineering **2017** (2017)
52. Brandao, P., Zisimopoulos, O., Mazomenos, E., Ciuti, G., Bernal, J., Visentini-Scarzanella, M., Menciassi, A., Dario, P., Koulaouzidis, A., Arezzo, A., et al.: Towards a computed-aided diagnosis system in colonoscopy: Automatic polyp segmentation using convolution neural networks. Journal of Medical Robotics Research **3**(02), 1840,002 (2018)
53. Laina, I., Rieke, N., Rupprecht, C., Vizcaíno, J.P., Eslami, A., Tombari, F., Navab, N.: Concurrent segmentation and localization for tracking of surgical instruments. In: International Conference on Medical Image Computing and Computer-Assisted Intervention, pp. 664–672. Springer (2017)
54. Attia, M., Hossny, M., Nahavandi, S., Asadi, H.: Surgical tool segmentation using a hybrid deep CNN-RNN auto encoder-decoder. In: IEEE International Conference on Systems, Man, and Cybernetics, pp. 3373–3378. IEEE (2017)
55. García-Peraza-Herrera, L.C., Li, W., Gruijthuijsen, C., Devreker, A., Attilakos, G., Deprest, J., Vander Poorten, E., Stoyanov, D., Vercauteren, T., Ourselin, S.: Real-time segmentation of non-rigid surgical tools based on deep learning and tracking. In: International Workshop on Computer-Assisted and Robotic Endoscopy, pp. 84–95. Springer (2016)
56. Milletari, F., Rieke, N., Baust, M., Esposito, M., Navab, N.: CFCM: Segmentation via coarse to fine context memory. In: A.F. Frangi, J.A. Schnabel, C. Davatzikos, C. Alberola-López, G. Fichtinger (eds.) Medical Image Computing and Computer Assisted Intervention, pp. 667–674. Springer International Publishing, Cham (2018)
57. Yu, L., Chen, H., Dou, Q., Qin, J., Heng, P.A.: Integrating online and offline three-dimensional deep learning for automated polyp detection in colonoscopy videos. IEEE Journal of Biomedical and Health Informatics **21**(1), 65–75 (2017)
58. Yu, F., Koltun, V.: Multi-scale context aggregation by dilated convolutions. arXiv preprint arXiv:1511.07122 (2015)
59. Bozkurt, A., Kose, K., Alessi-Fox, C., Gill, M., Dy, J., Brooks, D., Rajadhyaksha, M.: A multiresolution convolutional neural network with partial label training for annotating reflectance confocal microscopy images of skin. In: A.F. Frangi, J.A. Schnabel, C. Davatzikos, C. Alberola-López, G. Fichtinger (eds.) Medical Image Computing and Computer Assisted Intervention, pp. 292–299. Springer International Publishing, Cham (2018)
60. Springenberg, J.T., Dosovitskiy, A., Brox, T., Riedmiller, M.: Striving for simplicity: The all convolutional net. arXiv preprint arXiv:1412.6806 (2014)
61. He, J.Y., Wu, X., Jiang, Y.G., Peng, Q., Jain, R.: Hookworm detection in wireless capsule endoscopy images with deep learning. IEEE Transactions on Image Processing **27**(5), 2379–2392 (2018)
62. Mahmood, F., Durr, N.J.: Deep learning and conditional random fields-based depth estimation and topographical reconstruction from conventional endoscopy. Medical Image Analysis (2018)
63. Furukawa, R., Mizomori, M., Hiura, S., Oka, S., Tanaka, S., Kawasaki, H.: Wide-area shape reconstruction by 3D endoscopic system based on CNN decoding, shape registration and fusion. In: OR 2.0 Context-Aware Operating Theaters, Computer Assisted Robotic Endoscopy, Clinical Image-Based Procedures, and Skin Image Analysis, pp. 139–150. Springer (2018)
64. Oda, M., Roth, H.R., Kitasaka, T., Furukawa, K., Miyahara, R., Hirooka, Y., Goto, H., Navab, N., Mori, K.: Colon shape estimation method for colonoscope tracking using recurrent neural networks. In: A.F. Frangi, J.A. Schnabel, C. Davatzikos, C. Alberola-López, G. Fichtinger (eds.) Medical Image Computing and Computer Assisted Intervention, pp. 176–184. Springer International Publishing, Cham (2018)
65. Zhang, J., Xie, Y., Wu, Q., Xia, Y.: Skin lesion classification in dermoscopy images using synergic deep learning. In: International Conference on Medical Image Computing and Computer-Assisted Intervention, pp. 12–20. Springer (2018)
66. Lopez, A.R., Giro-i Nieto, X., Burdick, J., Marques, O.: Skin lesion classification from dermoscopic images using deep learning techniques. In: International Conference on Biomedical Engineering, pp. 49–54. IEEE (2017)

67. Navarro, F., Conjeti, S., Tombari, F., Navab, N.: Webly supervised learning for skin lesion classification. In: International Conference on Medical Image Computing and Computer-Assisted Intervention, pp. 398–406. Springer (2018)
68. Pal, A., Chaturvedi, A., Garain, U., Chandra, A., Chatterjee, R.: Severity grading of psoriatic plaques using deep CNN based multi-task learning. In: International Conference on Pattern Recognition, pp. 1478–1483. IEEE (2016)
69. Mendel, R., Ebigbo, A., Probst, A., Messmann, H., Palm, C.: Barrett's esophagus analysis using convolutional neural networks. In: Bildverarbeitung für die Medizin 2017, pp. 80–85. Springer (2017)
70. Georgakopoulos, S.V., Iakovidis, D.K., Vasilakakis, M., Plagianakos, V.P., Koulaouzidis, A.: Weakly-supervised convolutional learning for detection of inflammatory gastrointestinal lesions. In: IEEE International Conference on Imaging Systems and Techniques, pp. 510–514. IEEE (2016)
71. Hong, J., Park, B.y., Park, H.: Convolutional neural network classifier for distinguishing barrett's esophagus and neoplasia endomicroscopy images. In: 39th Annual International Conference of the IEEE Engineering in Medicine and Biology Society, pp. 2892–2895. IEEE (2017)
72. Ribeiro, E., Uhl, A., Wimmer, G., Häfner, M.: Exploring deep learning and transfer learning for colonic polyp classification. Computational and Mathematical Methods in Medicine (2016)
73. Yuan, Y., Qin, W., Ibragimov, B., Han, B., Xing, L.: RIIS-DenseNet: Rotation-invariant and image similarity constrained densely connected convolutional network for polyp detection. In: International Conference on Medical Image Computing and Computer-Assisted Intervention, pp. 620–628. Springer (2018)
74. Aoki, T., Yamada, A., Aoyama, K., Saito, H., Tsuboi, A., Nakada, A., Niikura, R., Fujishiro, M., Oka, S., Ishihara, S., et al.: Automatic detection of erosions and ulcerations in wireless capsule endoscopy images based on a deep convolutional neural network. Gastrointestinal Endoscopy (In press)
75. Fan, S., Xu, L., Fan, Y., Wei, K., Li, L.: Computer-aided detection of small intestinal ulcer and erosion in wireless capsule endoscopy images. Physics in Medicine & Biology **63**(16), 165,001 (2018)
76. Sekuboyina, A.K., Devarakonda, S.T., Seelamantula, C.S.: A convolutional neural network approach for abnormality detection in wireless capsule endoscopy. In: IEEE International Symposium on Biomedical Imaging, pp. 1057–1060. IEEE (2017)
77. Seguí, S., Drozdzal, M., Pascual, G., Radeva, P., Malagelada, C., Azpiroz, F., Vitrià, J.: Generic feature learning for wireless capsule endoscopy analysis. Computers in Biology and Medicine **79**, 163–172 (2016)
78. Vasilakakis, M.D., Diamantis, D., Spyrou, E., Koulaouzidis, A., Iakovidis, D.K.: Weakly supervised multilabel classification for semantic interpretation of endoscopy video frames. Evolving Systems pp. 1–13 (2018)
79. Itoh, T., Kawahira, H., Nakashima, H., Yata, N.: Deep learning analyzes helicobacter pylori infection by upper gastrointestinal endoscopy images. Endoscopy International Open **6**(2), E139 (2018)
80. Yu, J.s., Chen, J., Xiang, Z., Zou, Y.X.: A hybrid convolutional neural networks with extreme learning machine for WCE image classification. In: IEEE International Conference on Robotics and Biomimetics, pp. 1822–1827. IEEE (2015)
81. Chen, H., Wu, X., Tao, G., Peng, Q.: Automatic content understanding with cascaded spatial–temporal deep framework for capsule endoscopy videos. Neurocomputing **229**, 77–87 (2017)
82. Aubreville, M., Knipfer, C., Oetter, N., Jaremenko, C., Rodner, E., Denzler, J., Bohr, C., Neumann, H., Stelzle, F., Maier, A.: Automatic classification of cancerous tissue in laserendomicroscopy images of the oral cavity using deep learning. Scientific Reports **7**(1), 11,979 (2017)
83. Xu, T., Zhang, H., Huang, X., Zhang, S., Metaxas, D.N.: Multimodal deep learning for cervical dysplasia diagnosis. In: International Conference on Medical Image Computing and Computer-Assisted Intervention, pp. 115–123. Springer (2016)

84. Zou, Y., Li, L., Wang, Y., Yu, J., Li, Y., Deng, W.: Classifying digestive organs in wireless capsule endoscopy images based on deep convolutional neural network. In: IEEE International Conference on Digital Signal Processing, pp. 1274–1278. IEEE (2015)
85. Zhou, T., Han, G., Li, B.N., Lin, Z., Ciaccio, E.J., Green, P.H., Qin, J.: Quantitative analysis of patients with celiac disease by video capsule endoscopy: A deep learning method. Computers in Biology and Medicine **85**, 1–6 (2017)
86. Du, X., Kurmann, T., Chang, P.L., Allan, M., Ourselin, S., Sznitman, R., Kelly, J.D., Stoyanov, D.: Articulated multi-instrument 2D pose estimation using fully convolutional networks. IEEE Transactions on Medical Imaging (2018)
87. Colleoni, E., Moccia, S., Du, X., De Momi, E., Stoyanov, D.: Deep learning based robotic tool detection and articulation estimation with spatio-temporal layers. IEEE Robotics and Automation Letters **4**(3), 2714–2721 (2019)
88. Aubreville, M., Stoeve, M., Oetter, N., Goncalves, M., Knipfer, C., Neumann, H., Bohr, C., Stelzle, F., Maier, A.: Deep learning-based detection of motion artifacts in probe-based confocal laser endomicroscopy images. International Journal of Computer Assisted Radiology and Surgery (In press)
89. Sano, Y., Emura, F., Ikematsu, H.: Narrow-band imaging. Colonoscopy: Principles and Practice, Second Edition pp. 514–526 (2009)
90. Li, Q., He, X., Wang, Y., Liu, H., Xu, D., Guo, F.: Review of spectral imaging technology in biomedical engineering: achievements and challenges. Journal of Biomedical Optics **18**(10), 100,901–100,901 (2013)
91. Zeng, C., King, D.J., Richardson, M., Shan, B.: Fusion of multispectral imagery and spectrometer data in UAV remote sensing. Remote Sensing **9**(7), 696 (2017)
92. Wirkert, S.J., Vemuri, A.S., Kenngott, H.G., Moccia, S., Götz, M., Mayer, B.F., Maier-Hein, K.H., Elson, D.S., Maier-Hein, L.: Physiological parameter estimation from multispectral images unleashed. In: International Conference on Medical Image Computing and Computer-Assisted Intervention, pp. 134–141. Springer (2017)
93. Moccia, S., Vanone, G.O., De Momi, E., Laborai, A., Guastini, L., Peretti, G., Mattos, L.S.: Learning-based classification of informative laryngoscopic frames. Computer Methods and Programs in Biomedicine **158**, 21–30 (2018)
94. McLaren, B., Ashley, K.: Helping a CBR program know what it knows. Case-Based Reasoning Research and Development pp. 377–391 (2001)
95. Obermeyer, Z., Emanuel, E.J.: Predicting the futurebig data, machine learning, and clinical medicine. The New England journal of medicine **375**(13), 1216 (2016)
96. Pereira, F., Mitchell, T., Botvinick, M.: Machine learning classifiers and fmri: a tutorial overview. Neuroimage **45**(1), S199–S209 (2009)
97. Zhang, Q., Wu, Y.N., Zhu, S.: Interpretable convolutional neural networks. CoRR **abs/1710.00935** (2017). URL http://arxiv.org/abs/1710.00935
98. Wang, W., Shen, J.: Deep visual attention prediction. IEEE Transactions on Image Processing **27**(5), 2368–2378 (2018)
99. Sweeney, L.: k-anonymity: A model for protecting privacy. International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems **10**(05), 557–570 (2002)
100. Cocos, A., Qian, T., Callison-Burch, C., Masino, A.J.: Crowd control: Effectively utilizing unscreened crowd workers for biomedical data annotation. Journal of Biomedical Informatics **69**, 86–92 (2017)
101. Maier-Hein, L., Ross, T., Gröhl, J., Glocker, B., Bodenstedt, S., Stock, C., Heim, E., Götz, M., Wirkert, S., Kenngott, H., et al.: Crowd-algorithm collaboration for large-scale endoscopic image annotation with confidence. In: International Conference on Medical Image Computing and Computer-Assisted Intervention, pp. 616–623. Springer (2016)
102. Ross, T., Zimmerer, D., Vemuri, A., Isensee, F., Wiesenfarth, M., Bodenstedt, S., Both, F., Kessler, P., Wagner, M., Müller, B., et al.: Exploiting the potential of unlabeled endoscopic video data with self-supervised learning. International Journal of Computer Assisted Radiology and Surgery pp. 1–9 (2018)

103. Reinke, A., Eisenmann, M., Onogur, S., Stankovic, M., Scholz, P., Full, P.M., Bogunovic, H., Landman, B.A., Maier, O., Menze, B., et al.: How to exploit weaknesses in biomedical challenge design and organization. In: International Conference on Medical Image Computing and Computer-Assisted Intervention, pp. 388–395. Springer (2018)
104. Moccia, S., Mattos, L.S., Patrini, I., Ruperti, M., Poté, N., Dondero, F., Cauchy, F., Sepulveda, A., Soubrane, O., De Momi, E., et al.: Computer-assisted liver graft steatosis assessment via learning-based texture analysis. International Journal of Computer Assisted Radiology and Surgery pp. 1–11 (2018)
105. Bernardini, M., Romeo, L., Misericordia, P., Frontoni, E.: Discovering the type 2 diabetes in electronic health records using the sparse balanced support vector machine. IEEE Journal of Biomedical and Health Informatics (2019)
106. Ganganwar, V.: An overview of classification algorithms for imbalanced datasets **2**, 42–47 (2012)
107. Heikkilä, T., Dalgaard, L., Koskinen, J.: Designing autonomous robot systems-evaluation of the r3-cop decision support system approach. In: SAFECOMP 2013-Workshop DECS (ERCIM/EWICS Workshop on Dependable Embedded and Cyber-physical Systems) of the 32nd International Conference on Computer Safety, Reliability and Security, p. NA (2013)
108. Hansen, P., Ombler, F.: A new method for scoring additive multi-attribute value models using pairwise rankings of alternatives. Journal of Multi-Criteria Decision Analysis **15**(3-4), 87–107 (2008)
109. D'Haese, P.F., Konrad, P.E., Pallavaram, S., Li, R., Prassad, P., Rodriguez, W., Dawant, B.M.: CranialCloud: a cloud-based architecture to support trans-institutional collaborative efforts in neurodegenerative disorders. International Journal of Computer Assisted Radiology and Surgery **10**(6), 815–823 (2015)