LapGyn4: A Dataset for 4 Automatic Content Analysis Problems in the Domain of Laparoscopic Gynecology

Andreas Leibetseder, Stefan Petscharnig, Manfred Jürgen Primus, Sabrina Kletz, Bernd Münzer, Klaus Schoeffmann
Institute of Information Technology
Alpen-Adria University, 9020 Klagenfurt Austria
{aleibets|spetsch|mprimus|skletz|bernd|ks}@itec.aau.at

Jörg Keckstein
University Ulm,
Medizinische Fakultät
Ulm, Germany
joerg@keckstein.at

ABSTRACT

Modern imaging technology enables medical practitioners to perform minimally invasive surgery (MIS), i.e. a variety of medical interventions inflicting minimal trauma upon patients, hence, greatly improving their recoveries. Not only patients but also surgeons can benefit from this technology, as recorded media can be utilized for speeding-up tedious and time-consuming tasks such as treatment planning or case documentation. In order to improve the predominantly manually conducted process of analyzing said media, with this work we publish four datasets extracted from gynecologic, laparoscopic interventions with the intend on encouraging research in the field of post-surgical automatic media analysis. These datasets are designed with the following use cases in mind: medical image retrieval based on a query image, detection of instrument counts, surgical actions and anatomical structures, as well as distinguishing on which anatomical structure a certain action is performed. Furthermore, we provide suggestions for evaluation metrics and first baseline experiments.

CCS CONCEPTS

• **Information systems** → **Test collections**; • **Applied computing** → *Health informatics*; • **Computing methodologies** → *Neural networks*;

KEYWORDS

medical endoscopy data sets, surgical actions, instruments, anatomy

ACM Reference Format:
Andreas Leibetseder, Stefan Petscharnig, Manfred Jürgen Primus, Sabrina Kletz, Bernd Münzer, Klaus Schoeffmann and Jörg Keckstein. 2018. LapGyn4: A Dataset for 4 Automatic Content Analysis Problems in the Domain of Laparoscopic Gynecology. In *MMSys'18: 9th ACM Multimedia Systems Conference, June 12–15, 2018, Amsterdam, Netherlands.* ACM, New York, NY, USA, 6 pages. https://doi.org/10.1145/3204949.3208127

MMSys'18, June 12–15, 2018, Amsterdam, Netherlands

ACM ISBN 978-1-4503-5192-8/18/06…$15.00
https://doi.org/10.1145/3204949.3208127

1 INTRODUCTION

Medical procedures carried out via inserting small instruments into the human body aptly are referred to as minimally invasive surgeries (MISs), since, in contrast to traditional surgeries where inner bodily regions of patients are laid open, all operative equipment is introduced in a much less invasive manner, i.e. through natural or artificially created orifices of smallest conceivable magnitudes. Such practices greatly speed up a patient's healing process, which is why today many surgeries such as *cholecystectomy* almost exclusively are performed via the most popular form of MIS – *laparoscopy* [6].

Because operating physicians during endoscopic interventions utterly rely on externally positioned monitors displaying live video feeds that are transmitted by similarly introduced small cameras (*endoscopes*), these devices are required to deliver a clear, illuminated signal and provide enough flexibility for covering a broad field of view. With such optical technology readily available, it comes as no surprise that ever more hospitals record entire surgeries and put them into long-term storage – at least for the purpose of legal safeguarding. Being able to recall every step of every conducted procedure not only provides lawful protection, the material can as well be consulted for important tasks physicians perform on a daily basis: briefing patients, planning treatments and creating case documentation.

From a physician's point of view fully documented surgeries, on one hand, are a blessing since they simply contain all information about corresponding interventions, yet, on the other hand, can as well be considered a curse, because unless properly preprocessed their manual perusal is tiresome and time-consuming. Therefore, with the advent of more and more endoscopic media being recorded comes the incentive for various scientific communities to conduct research on suitably processing this kind of data (semi-)automatically [3].

Since the field of post-procedural analysis of endoscopic videos is fairly new and as of yet surprisingly rather scarcely researched, despite the abundance of hospitals recording surgeries, gathering enough media for research can be a challenging endeavor as such data is widely considered highly sensitive regarding patient privacy and therefore more often than not heavily protected by law. Hence, currently researchers regularly resort to establishing individual collaborations with medical facilities in order to conduct research on MIS data that frequently can not be published legally, which not only renders their results irreproducible but as well greatly hinders comparability to other researchers' work evaluated on different datasets. For this reason as well as for encouraging research in the domain of endoscopic media analysis, we purposefully introduce LapGyn4[1] – a four-part *gynecologic laparoscopy*[2] dataset comprising collections of images depicting general surgical actions, anatomical structures, conducted actions on specific anatomy as well as examples of differing amounts of visible instruments. Mainly targeted at image retrieval and classification but also suitable for other tasks such as segmentation, the dataset, although restricted to the specific endoscopic sub-domain of laparoscopy, can be utilized for a great variety of applications such as detecting visible anatomy or instruments, finding similarities between surgical actions and pointing out relevant scenes by recognizing a certain action on a particular body part. Within the following sections a detailed description of LapGyn4's components is provided (Section 2), several evaluation methodologies are proposed (Section 3) and results of early baseline analyses using convolutional neural network (CNN) GoogLeNet [5] are discussed (Section 4).

2 DATASET DESCRIPTION

This section describes the individual datasets as well as the applied collection methodology. Surgical Actions (Section 2.1) and Anatomical Structures (Section 2.2) have been extracted from 111 different gynecologic laparoscopic interventions, while Actions on Anatomy (Section 2.3) and Instrument Count (Section 2.4) were taken from 411 individual surgeries. Additionally, for Instrument Count the publicly available Cholec80 dataset [7] has been included. Table 1 offers an overview over all four datasets listing the sample count of their individual image classes, which are visually as well as textually described in following subsections that furthermore point out potential use cases.

Surgical Actions		**Anatomical Structures**	
Suction & Irrigation	3036	Uterus	938
Suturing	12914	Ovary	1162
Cutting (C)	1185	Oviduct	195
Cutting (HF)	3752	Liver	138
Dissection (blunt)	1444	Colon	295
Sling (Hyst.)	2752		
Coagulat.	3480		
Injection	2119		
Total	**30682**		**2728**
Actions on Anatomy		**Instrument Count**	
Suturing (Uterus)	232	0 Instr.	5100
Suturing (Ovary)	196	1 Instr.	5206
Suturing (Vagina)	263	2 Instr.	5856
Suturing (Other)	338	3 Instr.	5271
Total	**1029**		**21433**

Table 1: Datasets – classes and image counts

2.1 Surgical Actions

Surgical Actions are general activities performed during surgery involving one or more instruments. A computer-aided distinction between them in post-surgical analyses could certainly pose an advantage for physicians, since they represent the most critical component in medical procedures and, therefore, can be considered the primary target when it comes to seeking for a specific surgical phases. The dataset altogether contains over 30K images.

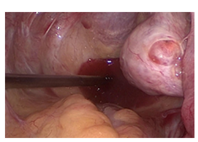

Fig. 1: Suction & Irrigation

2.1.1 Suction & Irrigation. These type of images depict the utilization of a suction and irrigation tube, which features a variety of purposes. Irrigation is the action of cleaning tissue from blood and other liquids in order to provide operating surgeons with clear visibility on corresponding bodily regions. For this action the tube sprays a ray of liquid on the tissue to be cleaned. The introduced liquid certainly at some point needs to be evacuated, hence, the tube offers a suction functionality for removing liquids from affected regions. Since this instrument is often additionally used for positioning as well as palpating tissue misclassifications can arise when merely considering the tube itself as a criterion for suction and irrigation.

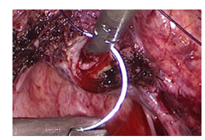

Fig. 2: Suturing

2.1.2 Suturing. Suturing is the action of stitching together tissue. Therefore, corresponding images mainly show the equipment used to accomplish this task: surgical needle, tissue as well as needle holder, knot pusher and suturing thread. The shape of surgical needles can be rounded or straight and they often are partially or even fully occluded. The suture itself varies in color, thickness and as well in type, since a variety of suturing techniques fitting various purposes are applied in modern endoscopy.

Fig. 3: Cutting (C)

2.1.3 Cutting (cold). Cold (C) cutting in this dataset refers to separating tissue with non electrically heated sharp instruments, such as cold surgical scissors or scalpels. Typically these kind of images often not only contain just the utilized cutting instrument but additionally an instrument for fixating loose tissue, which also applies for all other actions in this dataset, concerned with separating tissue, such as high-frequency cutting as well as blunt dissection.

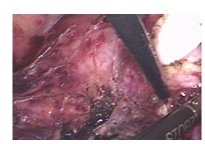

Fig. 4: Cutting (HF)

2.1.4 Cutting (high frequency). High frequency (HF) cutting indicates the action of tissue separation using electrically heated cutting instruments, such as monopolar needles or bipolar scissors. On occasions, the electrical heating causes the top of the instrument to light up, which can be regarded as a specific visual feature of this kind of images. Additionally, a low to medium amount of smoke is

[1]http://www.itec.aau.at/ftp/datasets/LapGyn4/ or https://doi.org/10.5281/zenodo.1219280
[2]Gynecologic laparoscopic endoscopy is concerned with female reproductive organs in the abdominal region.

produced when treating tissue in this manner, which can constitute a significant property for the task of image classification.

2.1.5 Dissection (blunt). In addition to being subject to cutting, tissue is as well often dissected using blunt instruments. This procedure may cause minor bleeding from dissecting small vessels, which is why blunt dissection is sometimes followed by coagulation in order to close open wounds. Furthermore, no specific instrument can be tied to this action, as surgeons use any suitable tool such as the suction and irrigation tube to perform blunt dissection.

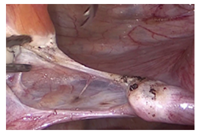

Fig. 5: Dissection (blunt)

2.1.6 Sling (Hysterectomy). This class of images show an important part of *hysterectomy* – the surgical procedure to remove a woman's uterus. The most prominent indicator for such an operation in laparoscopy is the visibility of an electrical sling that is wrapped around the lower end of corresponding uterus, i.e. the cervix, before using thermal dissection for removal. This action generally produces considerable amounts of smoke. Afterwards, often suturing and coagulation are performed in order to stitch together tissue as well as close open wounds.

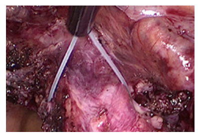

Fig. 6: Sling (Hyst.)

2.1.7 Coagulation. As many laparoscopic actions involve severing or cutting tissue, often causing bleeding wounds, electrically heated instruments in a process called coagulation are used to burn or *cauterize* these bodily regions. Depending on the severity of the wounds this procedure causes medium to high amounts of smoke to appear. Again, a variety of instruments such as graspers or scissors can be utilized for this kind of action, which introduces additional classification difficulties, since corresponding images could easily be misclassified as high frequency cutting or suturing scenes under certain circumstances.

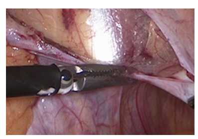

Fig. 7: Coagulation

2.1.8 Injection. During the injection procedure liquid is introduced into a patient's tissue for minimizing physical trauma. This is accomplished by inserting a straight rounded metal needle that, just as a suturing needle, can reflect light from the endoscope, hence, appear shiny. Additional instruments can be used in order to hold and guide the needle to a suitable injection spot. After completing the action the tissue around the tip of the needle usually inflates.

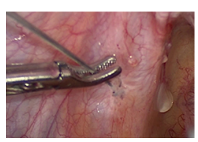

Fig. 8: Injection

2.2 Anatomical Structures

Anatomical Structures contains a collection of different inner bodily organs that are of significance for several laparoscopic procedures, such as the treatment of *endometriosis* – a benign but painful anomaly where uterine like tissue is found outside of the uterus [1]. In order to improve visibility for all involved classes, images mostly have been extracted from scenes where corresponding organs are inspected rather than operated on. This dataset is useful for disassociating organs from surgical actions, which allows for the retrieval of relevant scenes involving a particular anatomy. Altogether the dataset comprises 2.7K images.

2.2.1 Uterus. The uterus or womb is the part of the female reproductive system concerned with growing a fetus developed from a fertilized ovum. In laparoscopic surgeries the uterus often is operated on during myoma resection, i.e. the removal of tumors, or endometriosis treatment in the adenomyosis disease pattern, where the inner uterine tissue breaks through the wall of the uterus.

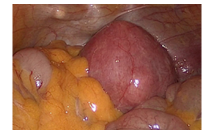

Fig. 9: Uterus

2.2.2 Ovary. Ovaries, besides other important functions such as producing hormones, have the purpose of developing mature ovums that later reside within the female uterus, where they are either fertilized or not. Surgical procedures involving ovaries often involve the removal of abnormal fluid accumulations, i.e. ovarian cysts. In case of endometriosis treatment ovaries are of importance for diagnosing adhesions.

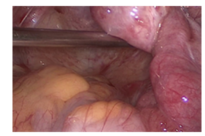

Fig. 10: Ovary

2.2.3 Oviduct. The oviducts or Fallopian tubes constitute the connection between Ovaries and Uterus, holding the main purpose of delivering matured ovums to the female uterus. Potential laparoscopic interventions affecting this organ include infertility correcting operations such as *fimbrioplasty*, drainage of cysts, or in the case of diagnosing endometriosis, similarly to the ovaries, the oviduct is as well often inspected for adhesions.

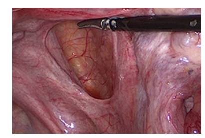

Fig. 11: Oviduct

2.2.4 Liver. The liver is part of the human digestive system: it among other functions is responsible for bile production, blood detoxification and excretion of a variety of substances. Since the organ is strongly exposed to a variety of toxins, concerning endoscopy in general, it is often subject for biopsies and in the commonly applied cholecystectomy involved in one of the most critical phases, where it is separated from the gall bladder. In laparoscopy the organ is treated for local cancer removal as well as for gynecologic endometriosis treatment.

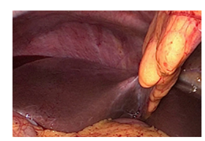

Fig. 12: Liver

2.2.5 Colon. The final anatomical structure is as well part of the human digestive system: the colon or large intestine as last part of the gastrointestinal tract stores bodily waste before the defecation process. Generally this organ is treated for abnormal growths such as polyps, *colorectal* cancer or colonic inertia. Cancer removal is performed laparoscopically and again concerning gynecology the colon is regarded for endometriosis treatment, as the condition is even more commonly diagnosed within the bowel region than on the liver.

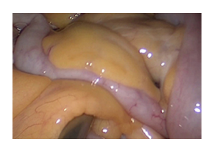

Fig. 13: Colon

2.3 Actions on Anatomy

This dataset comprises specific actions on particular anatomy and it at this point includes Suturing on the Uterus (Figure 14), an

Ovary (Figure 15), the Vagina (Figure 16) as well as Other regions (Figure 17).

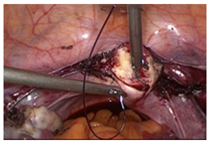

Fig. 14: Suturing (Uterus)

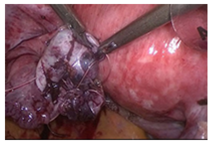

Fig. 15: Suturing (Ovary)

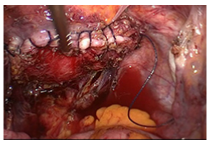

Fig. 16: Suturing (Vagina)

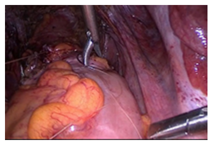

Fig. 17: Suturing (Other)

Although the combination of action and anatomy could potentially be identified using previous datasets, it makes sense to separately gather images of these particular scenes, since they can be relevant to post-procedural analyses of specific interventions. Over 1K images are contained in this dataset.

2.4 Instrument Count

The final dataset contains images of different amounts of surgical tools, in particular zero (Figure 18), one (Figure 19), two (Figure 20) and three (Figure 21) visible instruments.

Fig. 18: No instruments

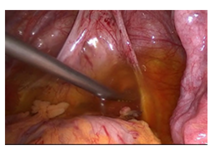

Fig. 19: One instrument

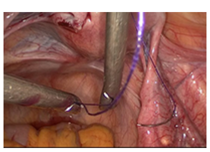

Fig. 20: Two instruments

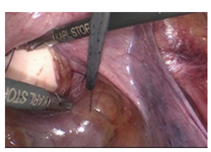

Fig. 21: Three instruments

Classifying the number of visible instruments may seem pointless at first glance, yet, it needs to be kept in mind that many surgical phases in endoscopy feature a specific amount of instruments – correctly identifying the amount of tools can, therefore, be a strong indicator for correctly identifying the corresponding phase during a procedure. Such an approach of course could be combined with identifying particular instrument types or even models in order to improve retrieval results, however, the sheer amount of different existing instruments and manufacturers makes it a challenging task to create a general purpose dataset. As an exemplary use case, this dataset can be used as basis for transfer learning experiments in the domain of cataract surgery, where the detection of idle phases (phases without instruments) is of great importance. The dataset contains over 21K images.

3 PROPOSED EVALUATION METHODOLOGY

In order to quantitatively evaluate each dataset for the task of image classification, we propose using 5-fold cross validation together with a suitable machine learning technique. Hence, we split every individual dataset into five subsets, which subsequently are utilized in the following manner: within each of altogether five iterations (folds) four subsets are used for training and one for testing with the limitation that every subset is required to be used for testing exactly once. In order to obtain an overall result, the results of each fold are averaged.

Regarding performance analyses, we follow the proposition made for KVASIR [4] – a public medical dataset for gastrointestinal disease detection – and suggest seven individual established metrics: Jaccard index (jacc), recall (rec), precision (prec), specificity (spec), accuracy (acc), Matthews correlation coefficient (mcc) and f1-value (f1), listed in Equations 1a-1g. All of these metrics are obtained by comparing classifier predictions to corresponding test data for following situations: a positively predicted class matches the actual class (true positive, TP), a negative prediction for a certain class is correct (true negative, TN), the predicted class does not match the actual class (false positive, FP) and a class has not been predicted as correctly (false negative, FN).

$$jacc = \frac{TP}{TP + FN + FP} \tag{1a}$$

$$rec = \frac{TP}{TP + FN} \tag{1b}$$

$$prec = \frac{TP}{TP + FP} \tag{1c}$$

$$spec = \frac{TN}{FP + TN} \tag{1d}$$

$$acc = \frac{TP + TN}{TP + TN + FP + FN} \tag{1e}$$

$$mcc = \frac{TP \times TN - FP \times FN}{\sqrt{(TP + FP)(TP + FN) + (TN + FP)(TN + FN)}} \tag{1f}$$

$$f1 = \frac{2TP}{2TP + FP + FN} \tag{1g}$$

While almost all metrics produce values within the interval of $[0, 1]$, with 1 being the optimal result, the mcc metric will result in the slightly different range of $[-1, 1]$, with -1 denoting total disagreement and 1 total agreement between predictions and observations. Furthermore, a value of 0 signifies that a classifier does not perform better than one that predicts randomly.

4 BASELINE EVALUATIONS

This section contains all preliminary evaluation results we achieve by utilizing an established and widely used CNN architecture. For all evaluations we employ the GoogLeNet [5] architecture in a learning from scratch fashion instead of using pre-trained models for transfer learning. This decision is based on the fact that standard models that have been trained on large image datasets, therefore, already capable of recognizing image characteristics can hardly constitute a basis for training on images from the endoscopic domain, since predominantly various key characteristics such as color composition or brightness as well as even shown objects greatly differ from images taken from everyday scenes. Model training is conducted on high-end machines equipped with Nvidia GeForce GTX 1080 Graphics cards and implemented with popular deep learning framework Caffe[3]. As we strive for a baseline evaluation we utilize standard parameters as well as the Adam optimization algorithm [2] throughout the learning process. Furthermore, training is conducted for 100 epochs retaining the best of them over all cross validation folds (cf. Section 3) for evaluations. In order to compensate for the imbalance of most datasets, we average retrieved class-specific performance metrics based on the corresponding sample count, i.e. calculate a weighted average (WAvg). The following sections outline the results obtained for each dataset, the contained tables of which indicate lowest values in *italic* and highest in **bold** letters.

4.1 Surgical Actions

Beginning with the largest of all datasets, Table 2 lists the evaluation results for Surgical actions, exhibiting notably high values on weighted average metrics – especially in correctly classifying that images do not belong to a certain class (cf. WAvg. spec=0.984).

	jacc	rec	prec	spec	acc	mcc	f1
Coagulation	0.719	0.821	0.852	0.982	*0.964*	0.816	0.837
Cutting (C)	*0.643*	*0.732*	0.841	0.994	0.984	*0.777*	*0.783*
Cutting (HF)	0.893	0.951	0.936	0.991	0.986	0.935	0.943
Dissection (blunt)	0.729	0.860	*0.827*	0.991	0.985	0.836	0.843
Injection	0.889	0.926	0.958	**0.997**	**0.992**	0.937	0.941
Sling (Hyst.)	0.892	0.954	0.932	0.993	0.990	0.937	0.943
Suction & Irrigation	0.749	0.862	0.851	0.983	0.971	0.841	0.857
Suturing	**0.947**	**0.978**	**0.968**	*0.976*	0.977	**0.953**	**0.973**
WAvg.	*0.864*	0.925	0.924	**0.984**	0.979	0.909	0.924

Table 2: Evaluations – Surgical Actions

This indicates that the classes themselves rather possess distinct features. Furthermore, suturing by far seems to be easiest to predict, as achieved performance for this class exceeds all other classes in every metric, except for spec/acc, the values of which peculiarly are slightly better in all other classes, yet, with ranges from 0.976-0.997 (spec) and 0.977-0.992 (acc) not at all far apart.

[3]http://caffe.berkeleyvision.org

4.2 Anatomical Structures

Judging the calculated metrics of this baseline evaluation with values of at least 0.800 throughout all classes, as Table 3 indicates, among all dataset evaluations this one achieves the best results for all contained categories, even surpassing the surprisingly good results for Surgical Actions. This could be due to the images nature: as mentioned most of them have been taken from scenes where the content is inspected in isolation, i.e. for the most part without other visible anatomy or even instruments.

	jacc	rec	prec	spec	acc	mcc	f1
Colon	0.836	0.919	*0.903*	0.988	0.981	**0.900**	0.911
Liver	0.810	*0.822*	**0.982**	**0.999**	**0.990**	0.894	0.895
Ovary	**0.862**	**0.945**	0.908	*0.929*	*0.936*	*0.870*	**0.926**
Oviduct	*0.800*	0.841	0.943	0.996	0.985	0.882	*0.889*
Uterus	0.854	0.917	0.925	0.961	0.946	0.880	0.921
WAvg.	*0.849*	0.919	0.920	**0.955**	0.951	0.879	0.918

Table 3: Evaluations – Anatomical Structures

Between all of the results liver seems to work best, which might be due to the organs distinct shape or color scheme involved: the liver appears dark red and is often surrounded by body fat or occasionally parts of the colon. Because of that fact, some misclassifications for colon could be explained, as indicated by the class often exhibiting the lowest values for many metrics.

4.3 Actions on Anatomy

As Table 4 illustrates, many classes in this dataset have been falsely classified since the overall FN and FP rates are high, i.e. the jacc is at most just over 0.507. This could very well be due to the dataset's current composition – as of yet it only comprises just over 1K images.

	jacc	rec	prec	spec	acc	mcc	f1
Other	0.452	0.603	0.643	0.836	*0.760*	0.447	0.622
Ovary	0.441	*0.513*	**0.758**	**0.961**	**0.875**	**0.555**	0.612
Uterus	*0.394*	0.639	*0.507*	*0.819*	0.778	*0.424*	*0.565*
Vagina	**0.507**	**0.704**	0.644	0.867	0.825	**0.555**	**0.673**
WAvg.	*0.451*	0.620	0.634	**0.864**	0.803	0.490	0.620

Table 4: Evaluations – Actions on Anatomy (Suturing)

Additionally when regarding the high specificity of at least 0.819, it seems much more correct negative classifications have been made rather than positive ones. Also attributable to the rather bad performance, it has to be considered, that in various cases the Other class, while, although in general depicting different organs, parts of the other organ class can as well be visible, since at times distinct shots are difficult to obtain.

4.4 Instrument Count

Finally, Table 5 demonstrates that the baseline classifier performs well on the Instrument Count dataset.

	jacc	rec	prec	spec	acc	mcc	f1
0 Instr.	**0.862**	**0.928**	**0.924**	**0.976**	**0.965**	**0.903**	**0.926**
1 Instr.	0.673	0.790	0.819	0.944	0.907	0.743	0.804
2 Instr.	*0.631*	*0.763*	*0.785*	*0.921*	*0.878*	*0.690*	*0.773*
3 Instr.	0.770	0.898	0.844	0.946	0.934	0.827	0.870
WAvg.	*0.730*	0.842	0.841	**0.946**	0.920	0.787	0.841

Table 5: Evaluations – Instrument Count

While altogether classifying zero instruments overall really outperforms all other classes with values of at least 0.862, the classifier seems to have difficulties for one or two instruments. Interestingly classification performance improves when considering the three instrument class. A baseline classifier, therefore, could at least be utilized for binary classification, i.e. distinguishing between visible instruments or no instruments.

When contemplating all achieved results using the proposed basic approach, it becomes clear that there is still room for classification performance improvements on all of LapGyn4's datasets, specifically when considering more critical applications in the medical domain that would require considerably higher classification accuracies. This of course is expected since these evaluations should only constitute an initial baseline for future research. Nevertheless, such results should certainly encourage interested researchers to develop more sophisticated approaches in order to advance in the domain of endoscopic multimedia retrieval.

5 CONCLUSION

With the intent on encouraging research in the field of laparoscopic gynecology as well as providing a means for obtaining reproducible results we present the LapGyn4 dataset comprising four datasets (Surgical Actions, Anatomical Structures, Actions on Anatomy, Instrument Count) of over 500 surgical interventions amounting to over 55K images. In a baseline evaluation for the task of image classification learning from scratch with a standard convolutional neural network we achieve medium to very good results, heavily depending on complexity of the image class in question. In future work, we plan on constantly extending the dataset, especially improving the comparatively small dataset for Actions on Anatomy by adding more examples as well as a greater variety of classes.

ACKNOWLEDGMENTS

This work is supported by Universität Klagenfurt and Lakeside Labs GmbH, Klagenfurt, Austria and funding from the European Regional Development Fund and the Carinthian Economic Promotion Fund (KWF) under grant KWF 20214 u. 3520/26336/38165.

REFERENCES

[1] J. Keckstein. 2017. Endometriosis in the Intestinal Tract – Important Facts for Diagnosis and Therapy. *coloproctology* 39, 2 (mar 2017), 121–133. https://doi.org/10.1007/s00053-017-0144-5

[2] Diederik P. Kingma and Jimmy Ba. 2014. Adam: A Method for Stochastic Optimization. *CoRR* abs/1412.6980 (2014). arXiv:1412.6980 http://arxiv.org/abs/1412.6980

[3] Bernd Münzer, Klaus Schoeffmann, and Laszlo Böszörmenyi. 2017. Content-based processing and analysis of endoscopic images and videos: A survey. *Multimedia Tools and Applications* (jan 2017). https://doi.org/10.1007/s11042-016-4219-z

[4] Konstantin Pogorelov, Kristin Ranheim Randel, Carsten Griwodz, Sigrun Losada Eskeland, Thomas de Lange, Dag Johansen, Concetto Spampinato, Duc-Tien Dang-Nguyen, Mathias Lux, Peter Thelin Schmidt, Michael Riegler, and Pål Halvorsen. 2017. KVASIR: A Multi-Class Image Dataset for Computer Aided Gastrointestinal Disease Detection. In *Proceedings of the 8th ACM on Multimedia Systems Conference (MMSys'17)*. ACM, New York, NY, USA, 164–169. https://doi.org/10.1145/3083187.3083212

[5] Christian Szegedy, Wei Liu, Yangqing Jia, Pierre Sermanet, Scott Reed, Dragomir Anguelov, Dumitru Erhan, Vincent Vanhoucke, and Andrew Rabinovich. 2015. Going deeper with convolutions. 1–9.

[6] Charlotte Tsui, Rachel Klein, and Matthew Garabrant. 2013. Minimally invasive surgery: national trends in adoption and future directions for hospital strategy. *Surgical Endoscopy* 27, 7 (jul 2013), 2253–2257. https://doi.org/10.1007/s00464-013-2973-9

[7] Andru P. Twinanda, Sherif Shehata, Didier Mutter, Jacques Marescaux, Michel de Mathelin, and Nicolas Padoy. 2017. EndoNet: A Deep Architecture for Recognition Tasks on Laparoscopic Videos. *IEEE Transactions on Medical Imaging* 36, 1 (jan 2017), 86–97. https://doi.org/10.1109/TMI.2016.2593957