

# Image-Based Smoke Detection in Laparoscopic Videos

Andreas Leibetseder<sup>(✉)</sup>, Manfred Jürgen Primus, Stefan Petscharnig,  
and Klaus Schoeffmann

Institute of Information Technology, Alpen-Adria University,  
9020 Klagenfurt, Austria  
{aleibets,mprimus,spetsch,ks}@itec.aau.at

**Abstract.** The development and improper removal of smoke during minimally invasive surgery (MIS) can considerably impede a patient's treatment, while additionally entailing serious deleterious health effects. Hence, state-of-the-art surgical procedures employ smoke evacuation systems, which often still are activated manually by the medical staff or less commonly operate automatically utilizing industrial, highly-specialized and operating room (OR) approved sensors. As an alternate approach, video analysis can be used to take on said detection process – a topic not yet much researched in aforementioned context. In order to advance in this sector, we propose utilizing an image-based smoke classification task on a pre-trained convolutional neural network (CNN). We provide a custom data set of over 30 000 laparoscopic smoke/non-smoke images, part of which served as training data for GoogLeNet-based [41] CNN models. To be able to compare our research for evaluation, we separately developed a non-CNN classifier based on observing the saturation channel of a sample picture in the HSV color space. While the deep learning approaches yield excellent results with Receiver Operating Characteristic (ROC) curves enclosing areas of over 0.98, the computationally much less costly analysis of an image's saturation histogram under certain circumstances can, surprisingly, as well be a good indicator for smoke with areas under the curves (AUCs) of around 0.92–0.97.

**Keywords:** Smoke detection · Endoscopy · Image processing · Deep learning

## 1 Introduction

Substantial advances in health care technology over the recent decades enabled minimally invasive surgery (MIS), i.e. medical operations inflicting as little as possible physical trauma upon patients, to become common practice in the clinical community. Nowadays, some surgical interventions almost exclusively are performed via MIS [46], such as the cholecystectomy procedure for attending gallbladder conditions. Regarding the technology applied in such or similar situations, physicians rely on video-monitoring their treatment of a patient's internal

anatomy – a *modus operandi* achievable by introducing a high definition camera or *endoscope* in addition to a variety of instruments through bodily orifices. The corresponding medical field, namely endoscopy, is sub-categorized by considering the insertion locality of said video device, which may be natural apertures such as nose (rhinoscopy), ear (otoscopy), anus (anoscopy) etc. or deliberately created incisions used in order to examine interior cavities of joints (arthroscopy), thorax (thoracoscopy) as well as of the most frequently inspected abdomen – a zone treatable via a broad number of procedures that comprise the field of *laparoscopy*, constituting the main focus of this study.

Many laparoscopic actions require severing tissue, which can create open wounds causing internal bleeding, a matter which usually needs to be tended to urgently. This typically is accomplished by suturing, i.e. sewing parts of the affected tissue back together and thereby helping natural hemostasis, as well as cauterization, that is using electrically heated or laser instruments<sup>1</sup> in order to mitigate or stop the hemorrhage. The latter either can be applied during dissection as to prevent aforementioned effects or afterwards in an attempt to seal afflicted regions. In any case, it is estimated that tissue cauterization is applied in well over 90% of all surgical procedures, generating yet another undesirable side-effect: a gaseous mixture consisting of 95% water and 5% chemical, biological as well as physical by-products [32] – materials comprising a *surgical smoke plume*. Potentially harmful contained substances like toxins, viruses or bacteria as well as ultrafine particulate matter renders exposure to such an entity a possibly serious health risk for both medical staff and patients, as is indicated in a great amount of scientific documents [5, 10, 14, 21, 34, 37, 43]. Thus, the necessity of removing surgical smoke swiftly and safely after its creation seems imperative in modern medicine, yet involved hazards still are underestimated, which can cause bad decisions like releasing corresponding fumes into the operating room (OR) air<sup>2</sup>, a not uncommon practice according to Sahaf et al. [5].

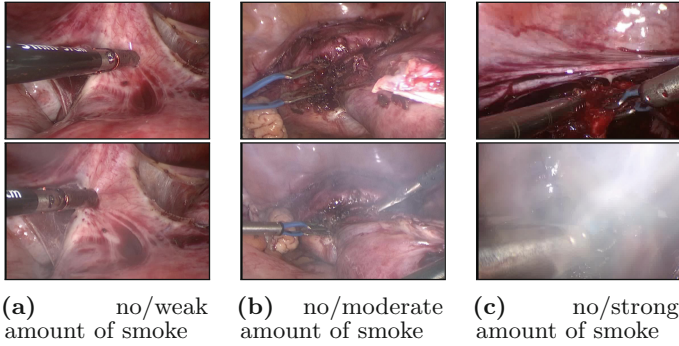
Proper smoke evacuation on the other hand is accomplished via OR-approved suction systems that typically are activated manually by the medical staff, in case cauterization is conducted. However, this particular action can easily be forgotten or neglected, potentially leading up to a point, in which the operating staff's view onto the currently treated body parts is severely obstructed by smoke – Fig. 1 demonstrates such situations by portraying three laparoscopic scenes that depict the emergence of smoke in various intensities.

In addition to the inconvenience of requiring manual control, smoke evacuators designed for laparoscopic utilization must be able to keep the abdominal cavity from collapsing during the suction process, which is achieved by using a medical grade insufflation gas<sup>3</sup> [7], entailing additional budget expenses to clinical institutions. Thus, handling a smoke evacuator inefficiently, which very likely happens many times during critical situations like surgeries, comes at a price. Naturally, automatic evacuation would represent an optimal solution for

<sup>1</sup> Temperatures range from about 100°–1200° Celsius.

<sup>2</sup> This effect is achieved by opening the stopcock of the laparoscopic port.

<sup>3</sup> In laparoscopy usually carbon dioxide ( $CO_2$ ) is used [35].



**Fig. 1.** Comparison of non-smoke vs. smoke images with different effect intensities.

both the nuisance of manual evacuator operation and the possibility of wasting valuable resources. Systems targeting similar goals have already been proposed, albeit all of them pursuing the rather naive methodology of commencing smoke removal whenever a cauterization instrument is activated [12, 13, 42]. Considering such a procedure fairly excessive and hardware restrictive, we argue that it is possible to construct more fine-grained, universal systems by detecting smoke via *image analysis* accurately and in real-time. Therefore, we formulate the research question behind our work as follows:

**Q** Can image-based analysis of endoscopic videos be leveraged as to reliably recognize the emergence of smoke in real-time?

Our proposed strategies to answer **Q** in general fall into the category of *binary classification tasks* – we develop a simple image saturation based histogram thresholding algorithm and compare its performance to two state-of-the-art CNN-based approaches.

The remainder of this work is subdivided into four sections: related work described in following Sect. 2, a detailed account of the methodology we apply in Sect. 3, evaluation results containing performance as well as runtime analyses in Sect. 4 and a concluding Sect. 5 highlighting our scientific contributions.

## 2 Related Work

Today classification utilizing CNNs is already commonly used in the medical field – research on the topic can be found dating back to the mid-1990s, where for example Sahiner et al. developed a three-layer CNN approach to be able to differentiate between normal tissue and abnormal areas (mass) when analyzing mammograms achieving a ROC AUC of 0.87 [40]. Further work using CNNs on computerized tomographic (CT) and Magnetic Resonance Imaging (MRI) images include Li et al. [30], who are detecting five different lung states related to interstitial lung diseases with 0.8 precision, 0.9 recall for each of them.

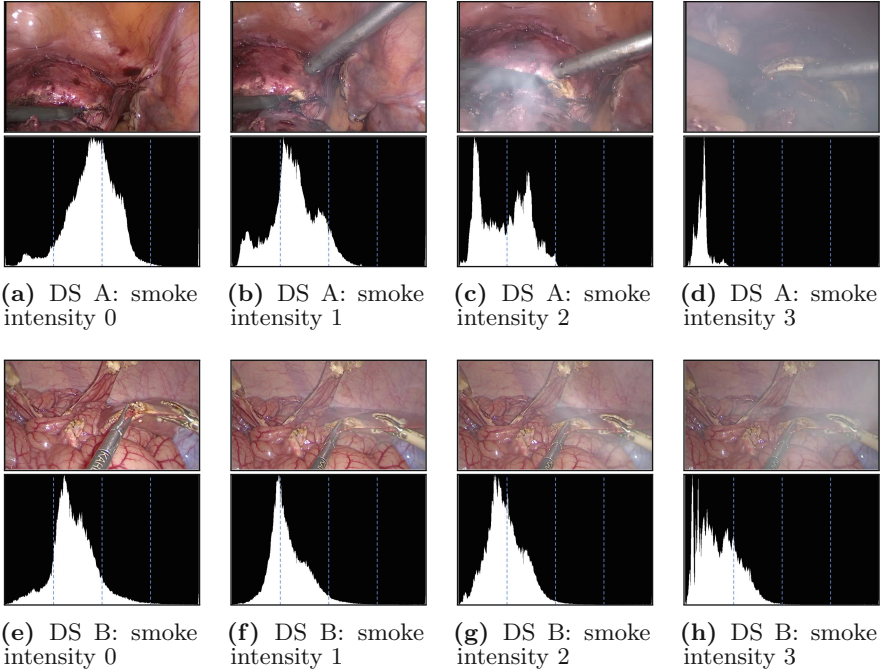
Conducting research in the same area, Anthimopoulos et al. [6] defined seven classes and they were able to outperform the former as well as other state-of-the-art methods. Moreover, Yan et al. [48] developed a multi-stage deep learning framework utilizing a CNN structure to automatically determine characteristics of different body parts, altogether exceeding recall, precision and F1 score of standard CNNs.

Although great potential for employing computer-aided processes in endoscopic surgery are being pointed out by Liedlgruber et al. [31], research concerned with classification techniques that operate on corresponding media yet is rather sparse – no matter if deep learning is applied or not. A few studies have been published by Häfner et al. within the scope of colonoscopy: they show the feasibility of automatically classifying colonic mucosa via feeding pyramidal discrete wavelet-transformed images to a k-nearest neighbors (k-NN) as well as Bayes classifier [17], develop a system for automated colon cancer detection based on the pit pattern classification (Kudo et al. [27]) in [18] and propose a novel color texture operator for pit pattern classification outperforming state-of-the-art operators in terms of compactness as well as computational speed [19]. As for CNN-based approaches, Park et al. [38] apply learning of hierarchical features on colonoscopy images for identifying polyp regions with an accuracy of 90%. Albeit in a different context, but specific to this work’s target-domain – laparoscopy – Petscharnig et al. [39] continue training AlexNet (Krizhevsky et al. [25]) to be able classify shots taken from a large gynecologic video database categorized into 14 different classes in order to aid physicians in the process of surgery annotation.

Finally, surgical smoke detection is yet another area still not much researched – predominantly visual smoke recognition is addressed in non-medical settings such as identifying fire outbursts [36,47,49], utilizing classification approaches like image separation [44], optical flow computation [11,24] or pattern recognition [15,16,45]. Since smoke emergence and lighting conditions in endoscopic environments strongly differ from outdoor settings, these techniques only to some extent are applicable to the medical sector. In the field of laparoscopy, apart from a non-vision-based assessment of smoke evacuation benefits (Takahashi et al. [42]) and an US patent from the Sony Corporation vaguely describing a frame-based system using motion blur as well as pixel block analysis [9], we merely are able to discover one related study, albeit targeted towards retrieval of scenes containing smoke in contrast to their real-time detection, as is our intent: Loukas et al. [33]. They extract 76 individual shots of 26–58 frames (between 1976–4408 images) from cholecystectomy videos, calculate their space-time optical flow together with some kinematic features and employ a one-class support vector machine (OCSVM) for classification, outperforming selected wavelet-based image decomposition methods for fire surveillance [8,16,29].

### 3 Proposed Methodologies

Altogether, we propose three smoke classification approaches: Sect. 3.1 gives an understanding of simply inspecting an image’s saturation channel in HSV color

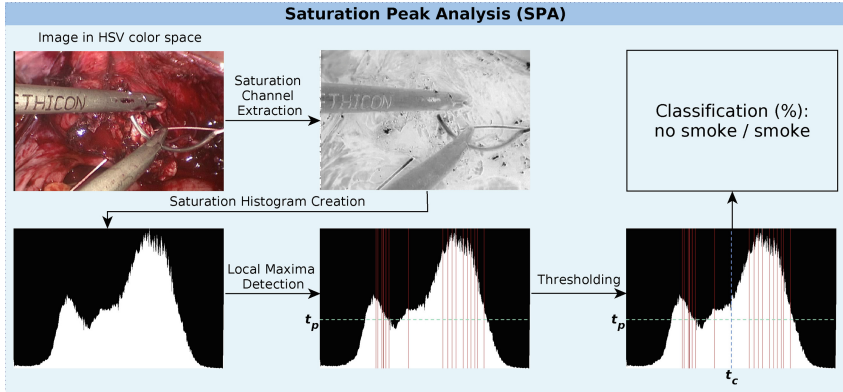


**Fig. 2.** Smoke development in different datasets DS A and B, including 256 bin saturation histograms. Images show various smoke intensities: none (0), weak (1), moderate (2), strong (3). Visual histogram comparison facilitated by division into four equal sectors (vertical lines).

space – a technique we call *Saturation Peak Analysis (SPA)* and Sect. 3.2 outlines the development of two GoogLeNet CNN models learned from both, full color (GLN RGB) as well as saturation only (GLN SAT) samples.

### 3.1 Saturation Peak Analysis (SPA)

Regions of smoke in endoscopic images tend to be grayish or rather *colorless*. Therefore, it seems appropriate to use the saturation component of the HSV color space to detect these areas, especially since the amount of smoke increases rapidly in the abdominal cavity when there is no evacuation mechanism in place. A caveat of taking such a perspective is that other colorless entities can be found during laparoscopic procedures: e.g. instruments and reflections of light hitting objects. Interferences like that can severely impact the saturation of an image, hence, naively observing this value will yield moderate classification results. Using the saturation histogram of a frame, we found in an explorative manner that by merely inspecting significant local bin maxima, i.e. *peaks* in the histogram’s shape, we can determine colorlessness, compensating for insignificant non-smoke influences.



**Fig. 3.** SPA Classification: finding local maxima in an image’s saturation histogram and classifying via thresholding. (Color figure online)

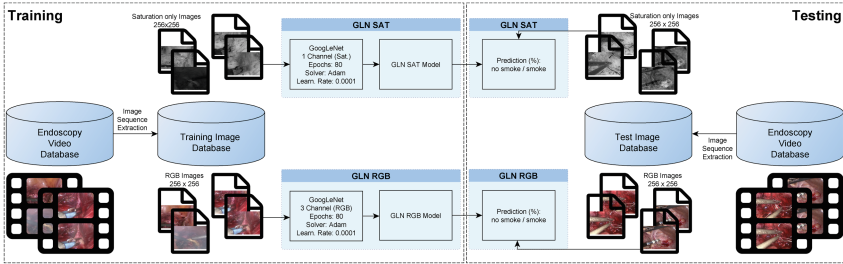
In order to illustrate the basis for our reasoning, Fig. 2 shows transitions in smoke intensities from no smoke to a very high degree of smoke together with corresponding saturation histograms for two scenes taken from different laparoscopic datasets<sup>4</sup>. Additionally to displaying individual pixel saturation counts via their 256 bins, the histogram images in the figure are sectioned into four equal parts indicated by three blue dashed vertical lines marking 25%, 50% and 75% portions of all bins, which helps facilitate their comparison across the portrayed smoke intensification. It can easily be discovered that the bin curves strongly correlate to the presence of smoke: for example, the depicted upper scene (Figs. 2a–d) starts out with an almost centered histogram curve (Fig. 2a) moving below the first bin quarter as smoke rises to a strong level (Fig. 2d). In contrast this development, the lower sequence’s histograms (Figs. 2e–h) overall are far less saturated, predominantly gathering in between the second bin portion (Fig. 2e) but swiftly gravitating below the first one at a high level of smoke (Fig. 2h), again indicating colorlessness in similar fashion to former example. Empirical pre-study analyses on our laparoscopic video material show that these individual trends apply to the majority of images in different datasets, therefore, smoke detection using saturation histograms seemingly boils down to finding an appropriate concentration point for bin values of non-smoke samples, i.e. a classification threshold as introduced shortly, which can be used as a reference to smoke samples that generally exhibit a lower concentration point. As this is not a straightforward task, at present we incrementally select such locations and apply SPA in order to classify a single image, which is visually described in Fig. 3.

SPA analyzes a frame’s saturation by converting it into the HSV color space, before isolating corresponding S-channel and creating a respective

<sup>4</sup> The image sequences show typical scenes from both of this study’s custom datasets, i.e. DS A and DS B (see Sect. 4.1 for details).

intensity histogram. Using this representation, a twofold decision criterium is employed, which in general relies on the above demonstrated observation that colorless/smoke-containing images exhibit many low saturated pixels, hence their corresponding histograms will comprise higher values in their lower bins, inherently establishing a vice versa situation for the upper ones (cf. Fig. 2). In detail, significant local maxima (peaks) are computed as a first step (red vertical solid lines in Fig. 3), restricted by the following iteratively determined constraints that as well constitute results of aforementioned empirical pre-study:

- A maximum must not be found below a *peak threshold* of  $t_p = 0.35 \times \text{max\_bin\_value}$  (green horizontal dashed line in Fig. 3), which ensures that a discovered peak is sufficiently significant.
- Left as well as right slopes culminating in a peak must be at least 2 bins wide rendering the peak's total width at least 5 bins, which eliminates small outliers exhibiting very similar saturation values (e.g. gray instruments).



**Fig. 4.** CNN Training/Testing: RGB/SAT images used in GoogLeNet-based model training, evaluations via different dataset.

Finally, classification is simply based on relating the number of peaks below a *classification threshold*  $t_c$  (blue vertical dashed line in Fig. 3) to the ones above, yielding prediction confidences  $pred_S$  for smoke as well as  $pred_{NS}$  for non-smoke, defined by Formulas 2 and 1:

$$pred_S(pk(H)) = \frac{|\{p \mid p \in pk(H) \wedge p \leq t_c\}|}{|pk(H)|}, \quad (1)$$

$$pred_{NS}(pk(H)) = \frac{|\{p \mid p \in pk(H) \wedge p > t_c\}|}{|pk(H)|}, \quad (2)$$

where  $H$  describes a set of input histogram bin values ( $|H| = 256$ ) and function  $pk(H) \subset \mathbb{N}_0$  calculates the set of peak positions following the criteria outlined above. In case no peak is found, i.e.  $pk(H) = \emptyset$ , the predictions are made via finding the majority of bin's values above and below  $t_c$ , defined by Formulas 4 and 3:

$$pred_S(H) = \frac{1}{|H|} \sum_{\substack{i=0 \\ b \in H \\ i \leq t_c}} b_i, \quad (3)$$



$$pred_{NS}(H) = \frac{1}{|H|} \sum_{\substack{i=0 \\ b \in H \\ i > t_c}} b_i. \quad (4)$$

For demonstration purposes, Fig. 3 indicates a  $t_c$  of 0.50, yet for evaluation values from 0.10 up to 0.80 in 0.05 increment steps are used, which, as mentioned, currently serves the purpose of iteratively finding suitable thresholds for videos exhibiting a different color spectrum. The necessity for this decision becomes apparent when recalling pre-study discovery, formerly highlighted when discussing Fig. 2: images from separate laparoscopic datasets on average show distinguishable differences in saturation histograms. Consequently, when once again regarding the illustrated smoke intensification examples, SPA should perform best between  $t_c = 0.40$  to  $t_c = 0.60$  for the first and  $t_c = 0.20$  to  $t_c = 0.40$  for the second scene, which will be evaluated in Sect. 4.

### 3.2 CNN Classification

Promising image classification results achieved by using CNN architectures, most prominently LeNet [28], AlexNet [26] and GoogLeNet [41] as well as advances in applying those networks in the medical domain (see Sect. 2) inspired our impulse to employ them for our smoke classification task at hand. While utilizing deeper networks like, for instance, ResNet [20] (152 layers) may yield better results, their slower computation speed would be detrimental to our general aim – real-time smoke detection on preferably commercially available hardware. Therefore, we choose to benefit from 22-layered pre-trained CNN architecture GoogLeNet and at first pursue the most conventional strategy of simply using RGB images to continue training the network, which we further denominate GLN RGB for brevity. In order to enable a direct comparison between a trained CNN model and the SPA approach that builds on saturation analysis, we use grayscale images only depicting the saturation channel of the HSV color space for creating a classification model we accordingly label GLN SAT – a decision largely based on discovering partially very promising results when applying SPA (see Sect. 4). Figure 4 illustrates both approaches for training and classification, which are conducted via popular deep learning framework Caffe [22].

For training and validating each of the GLN architectures an 80:20 split of dataset images<sup>5</sup> are used with an even distribution for non-smoke/smoke samples. Exclusively in case of GLN SAT these are converted to saturation only pictures, whereas further preprocessing remained the same for both methods: resizing to GoogLeNet’s intended resolution of  $256 \times 256$  pixels, computation of a global image mean needed for data normalization as well as encapsulating the results within a Lightning Memory-Mapped Database (lmdb) [2].

Model training altogether takes a little over two hours for each model on a machine running Linux Mint 17.3 (64-bit) [1] with following hardware specs:

<sup>5</sup> Approximately 20 000 non-smoke/smoke images of DS A (see Sect. 4.1 for details).



Intel Core i7-3770K CPU @ 3.50GHz x 4, 16 GiB DDR3 @ 1333 MHz, Nvidia GeForce GTX 980 Ti. The Caffe solver options have iteratively been adjusted through several training attempts and finally set to: 100 Epochs – ultimately we chose Epoch 80 due to its high accuracy, stochastic optimization using Adam [23] with an initial learning rate of 0.0001.

At last, classification can be conducted merely requiring the trained model (snapshot @ 80 Epochs) in order to calculate prediction confidences for non-smoke or smoke images.

## 4 Experimental Results

Detailed results of all three above described methodologies and statistics are covered within this section. First, we introduce our employed datasets in Sect. 4.1. Afterwards, a closer look is taken at evaluations using test data from DS A (Subsect. 4.2), which is taken from the same source material as the GLN training data, yet it of course comprises different scenes. Afterwards, images from DS B are evaluated, which, as already mentioned, are extracted from a distinctly separate kind of source (Sect. 4.3). Finally, the overall performance of the applied methods is inspected in Subsect. 4.4.

### 4.1 Datasets

All our evaluations are based on two datasets: dataset A (DS A) and dataset B (DS B), described in following short paragraphs.

DS A is used for training, validation as well as testing and it consists of images taken from over eight laparoscopic surgeries in the field of gynecology. We extract different frame sequences of up to two seconds in length, amounting to about 30 000 images, half of which show non-smoke situations, the other half depicts smoke occurring in various intensities. For training and validating CNN models we use approximately 20 000 images (50% non-smoke/smoke), which leaves about 10 000 samples for evaluations.

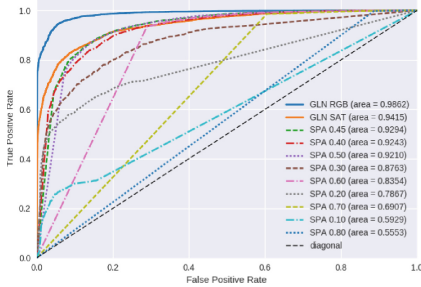
The laparoscopic source videos for DS A show many similarities, since they are recorded under similar conditions: the same endoscope and lighting yield an analogous image color spectrum. Therefore, we added DS B, which is extracted from a laparoscopic video recorded in another location and under different circumstances. The dataset’s color scheme differs in large parts from DS A, which we determined via a thorough preliminary histogram analysis and major implications, namely different optimal classification thresholds, are hinted at in Sect. 3.1, Fig. 2. Hence this dataset represents a valuable resource to solidify evaluation results. DS B consists of about 4 500 images (50% non-smoke/smoke), again taken from sequences of up to two seconds. They exclusively are used for evaluation only, which will be outlined in Sect. 4.3.

## 4.2 Evaluation Results - DS A

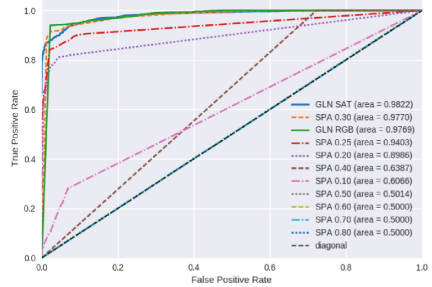
Results from evaluating DS A are illustrated in Table 1a, which lists selected classification measures for both GLN methods, as well as SPA with  $t_c$  ranging from 0.10 to 0.80 generally arranged in 0.10 increment steps except for exception  $t_c = 0.45$  in order to highlight its peak performance area (see details below). Classifications in the table are conducted at confidence  $c_c = 0.50$ , meaning for instance that in order to correctly classify an image containing smoke, the classifier’s prediction confidence for corresponding label needs to be 50% or higher (progression at different  $c_c$  values can be observed inspecting the ROC curve in Fig. 5a). For the given DS A, GLN RGB shows the best performance with 93.2% correctly classified smoke samples, i.e. very high sensitivity, and even higher specificity of 95.3%, i.e. correctly classified non-smoke samples, yielding an accuracy of 94.2%. GLN SAT achieves a slightly worse outcome but still yields a quite high accuracy of 87.0% with 82.6% sensitivity and 91.4% specificity. As for SPA, at  $c_c = 50$  a threshold of  $t_c = 0.40$  seems to classify similarly compared to GLN SAT, resulting in an accuracy of 85.0%, 87.7% sensitivity and 82.2% specificity. Regarding the accuracy and precision of SPA from  $t_c = 0.10$  up to  $t_c = 0.80$  it becomes clear that SPAs peak performance is around  $t_c = 0.30$  to  $t_c = 0.50$ , specifically above  $t_c = 0.40$ , which indicates that non-smoke saturation

**Table 1.** Evaluation results for datasets A and B,  $c_c = 0.50$ .

(a) DS A.							(b) DS B.							
Method	Accuracy	Precision	Sensitivity	Specificity	F1		Method	Accuracy	Precision	Sensitivity	Specificity	F1		
GLN RGB	0.942	0.952	0.932	0.953	0.942		GLN RGB	0.779	0.697	0.998	0.555	0.821		
GLN SAT	0.870	0.906	0.826	0.914	0.864		GLN SAT	0.914	0.879	0.962	0.864	0.919		
SPA 0.10	0.536	0.891	0.081	0.990	0.149		SPA 0.10	0.507	0.954	0.029	0.999	0.056		
SPA 0.20	0.674	0.968	0.360	0.988	0.525		SPA 0.20	0.815	0.994	0.639	0.996	0.778		
SPA 0.30	0.770	0.928	0.585	0.955	0.718		SPA 0.25	0.910	0.976	0.843	0.979	0.905		
SPA 0.40	0.850	0.831	0.877	0.822	0.854		SPA 0.30	0.892	0.843	0.966	0.816	0.900		
SPA 0.45	0.820	0.752	0.956	0.685	0.842		SPA 0.40	0.508	0.507	1.000	0.003	0.673		
SPA 0.50	0.738	0.659	0.986	0.491	0.790		SPA 0.50	0.507	0.507	1.000	0.000	0.672		
SPA 0.60	0.557	0.530	0.999	0.115	0.693		SPA 0.60	0.507	0.507	1.000	0.000	0.672		
SPA 0.70	0.506	0.503	1.000	0.013	0.670		SPA 0.70	0.507	0.507	1.000	0.000	0.672		
SPA 0.80	0.500	0.500	1.000	0.000	0.667		SPA 0.80	0.507	0.507	1.000	0.000	0.672		



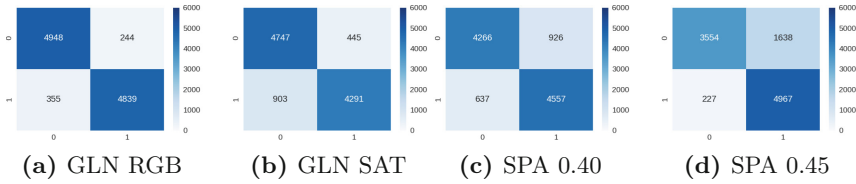
(a) DS A.



(b) DS B.

**Fig. 5.** ROC curve comparison for datasets A and B. (Color figure online)

histograms tend to exhibit more peaks, i.e. higher bin values, above  $t_c = 0.40$  and vice-versa for smoke histograms. Figure 6 shows the most significant confusion matrices at  $c_c = 0.50$ , used to calculate part of the results in Table 1a.



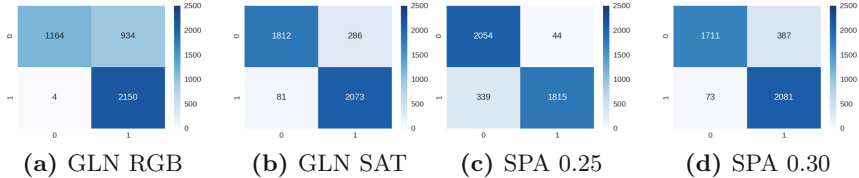
**Fig. 6.** Most significant confusion matrices for DS A (0 no smoke, 1 smoke),  $c_c = 0.50$ .

Clearly GLN RGB (Fig. 6a) with merely 599 misclassifications out of 10386 images again emphasizes the findings from above, whereas SPA 0.45 with 1865 (Fig. 6d) falsely classified samples stands out as the worst of the bunch. However, a slightly different impression can be gained when regarding a continuous  $c_c$  progression, as is depicted in Fig. 5a showing the ROC curve of the methods listed in Table 1a. Judging by the AUCs, it is evident that GLN RGB (solid blue curve) still performs best with an AUC of 0.9862, followed by GLN SAT’s (solid orange curve) AUC of 0.9415. For SPA although in contrast to the above discoveries  $t_c = 0.45$  (dashed green curve) seems to have an overall better performance than  $t_c = 0.40$  (dashed red curve), albeit just slightly (AUC 0.9294 vs. 0.9243). Nevertheless this is interesting to see, since results for  $c_c = 0.50$  seem to differ by a higher degree, which apparently is approximated as  $c_c$  progresses. SPA using other  $t_c$  values, as already pointed out, gradually perform worse up until the point of near randomness (dashed black diagonal line).

### 4.3 Evaluation Results - DS B

Due to the fact that DS B (around 4 000 images, 50% non-smoke/smoke), as mentioned above, has not been involved in any GLN training at all, it perfectly serves the purpose of further verifying previous findings. Its most salient difference to DS A has already been pointed out – a more or less consistently divergent color spectrum comprising much less saturated images. Therefore, the optimal  $t_c$  should definitely be lower than for DS A, which indeed is the case judging by the evaluation results at  $c_c = 0.50$  listed in Table 1b. This time GLN SAT seems to perform best yielding 91.4% classification accuracy, 96.2% sensitivity and 86.4% specificity. It is closely followed by SPA with  $t_c = 0.25$ , which as well achieves 91.0% accuracy but with almost interchanged sensitivity (84.3%) and specificity (97.9%) values, which indicates a better efficiency in detecting non-smoke than smoke. Nevertheless, the performance sweet spot for SPA seems to lie between  $t_c = 0.25$  and  $t_c = 0.30$ , since in the latter’s outcome sensitivity (96.6%) and specificity (81.6%) are again reversed, resulting in an accuracy of

89.2%. As Fig. 7 shows, GLN RGB at  $c_c = 0.50$  misclassifies a lot of non-smoke images (934 of 2098), which causes it to perform rather poorly compared to all other methods yielding unbalanced 100.0% sensitivity, 55.5% specificity and only 77.9% accuracy.



**Fig. 7.** Most significant confusion matrices for DS B (0 no smoke, 1 smoke),  $c_c = 0.50$ .

Finally, we take a look at the ROC curves from DS B’s evaluations, which are depicted in Fig. 5b and again paint a slightly different picture. GLN SAT (blue solid line) with an AUC of 0.9822 still turns out to be the best classifier for DS B. SPA with  $t_c = 0.30$  (orange dashed line), however, comes in second with an area of 0.9770, similarly to the DS A’s evaluation, outperforming the seemingly better SPA method at  $c_c = 0.50$ . Surprisingly GLN RGB (green solid line) ranks third with 0.9769 only performing negligibly worse than the former method. SPA with  $t_c = 0.25$  (red dashed line) classifies well yielding an AUC of 0.9403, yet performance for other SPA rapidly decreases, especially starting from  $t_c = 0.40$  upwards, where many effectively yield predictions equal to a random classifier – SPA curves above  $t_c = 60$  even exactly match the diagonal line.

#### 4.4 Runtime Evaluation

Since the intent behind this work is real-time smoke detection, it is important to as well consider computational performance in addition to above assessed classification quality. Table 2 shows the average wall clock timings<sup>6</sup> of image *preparation*, *classification* and their total for both datasets’ differing sample resolutions (DS A:  $720 \times 480$ , DS B:  $1920 \times 1080$ ).

**Table 2.** Image evaluation performance avg. in DS A/B (ms).

	Method	Preparation	Classification	Overall
	SPA (720x480)	3.542	0.005	3.546
	GLN RGB (720x480)	12.182	107.880	120.063
	GLN SAT (720x480)	6.234	75.936	82.170
	SPA (1920x1080)	12.487	0.006	12.493
	GLN RGB (1920x1080)	45.132	105.223	150.355
	GLN SAT (1920x1080)	18.847	75.307	94.154

<sup>6</sup> For the exact machine hardware specs, see Sect. 3.2.

All evaluations are implemented in Python [4] with preparation steps mostly consisting of OpenCV [3] tasks, like color conversion, image resizing and histogram extraction but as well of course a custom implementation for finding local maxima in case of SPA. Regarding the measurements for both resolutions, it becomes apparent that GLN RGB by far is the most costly of all methods with classification time requirements of about 105 ms, followed by GLN SAT with around 75 ms and SPA with negligible 0.005 ms. In case preparation timings are included, the overall processing duration worsens due to the relatively long time resizing images to  $256 \times 256$  pixels takes: depending on how many channels are used<sup>7</sup>, this step adds about 3–12 ms for  $720 \times 480$  and 8–45 ms for  $1920 \times 1080$ . This results in altogether 120–150 ms for GLN RGB, 82–94 ms for GLN SAT and 3–12 ms for SPA, rendering SPA the only method fulfilling real-time requirements<sup>8</sup> on the utilized test machine.

## 4.5 Discussion

When surveying the entirety of outcomes, a clear trend towards GoogLeNet using colored images (GLN RGB) can be observed, since its worst performance in both datasets still is producing a ROC AUC of above 0.97. Unfortunately this as well is the most computationally expensive method, showing runtime performances of about 150 ms per HD image, which indicates merely near real-time performance. Nevertheless, since smoke development across frames does generally not change very rapidly, it would very likely be feasible to drop some frames and still achieve great results in live systems. As an alternative, GoogLeNet fed with saturation images (GLN SAT) could be used to speed up the process considerably with a performance of around 94 ms for the same type of input. This would impact classification performance but not substantially, since at worst evaluations still show an AUC of over 0.94. The only method capable of true real-time performance is saturation peak analysis (SPA) with as little as around 12 ms computation requirements and ROC curve areas of at least over 0.92, when *always* considering the best classification threshold  $t_c$ . However, SPA critically relies on finding this right  $t_c$  for every classified image, which renders the algorithm, at least in its current form, inapplicable for live smoke detection. Still, when regarding analyses conducted on DS A and B, it seems apparent that, although different surgery setups can produce contrasting distributions in saturation, equivalent ones appear to share similar values. This consideration would for example explain SPA showing optimal performance for both datasets at different threshold ranges: around  $t_c = 0.40$  to  $t_c = 0.50$  for DS A and  $t_c = 0.20$  to  $t_c = 0.30$  for DS B.

Regarding comparability with most relevant work by Loukas et al. [33] described in Sect. 2, it has to be born in mind that the authors do not target real-time smoke evacuation, as is the case in our study. Nevertheless, since our methodologies can achieve at least a near real-time classification rate, they could

<sup>7</sup> SAT channel conversion takes around 3 ms for  $720 \times 480$ , 10 ms for  $1920 \times 1080$ .

<sup>8</sup> For a 25 fps video real-time requirements would be:  $\frac{1000}{25} = 40$  ms.

as well be utilized to annotate recorded media. In straight comparison, although outperforming selected wavelet-based outdoor smoke detection methods with an achieved ROC AUC of 0.63, their methodology seems to perform considerably worse than our proposed techniques, at least for their custom created dataset.

## 5 Conclusion

Targeting real-time smoke detection in endoscopic videos, we develop several image-based classification approaches, which we evaluate on two custom laparoscopic datasets. Continued training of GoogLeNet using full color samples overall achieves the highest classification but lowest runtime performance, which could be mitigated by simply omitting frames in real-time systems. Alternatively, using saturation channel only images for GoogLeNet training still produces a high accuracy at much faster computation times, yet as well not fully capable of handling live streams. In contrast to these CNN-based methods, naive image saturation analysis shows good performance in terms of classification and runtime, however, it is currently limited to requiring information about a dataset's average saturation distribution for non-smoke images.

When addressing our general research question **Q** inquiring the feasibility of reliable smoke recognition in laparoscopic live streams, we consider the achieved classification quality to be good enough for highly accurate systems. Regarding the real-time aspect, future investigations need to be conducted, although we estimate dropping frames being a sufficient measure to compensate for slower computation speeds. Furthermore, we deem the evaluated methodologies also be applicable to general endoscopic videos, since they typically are very similar to laparoscopic recordings, where equivalent equipment is used.

In future work, we will evaluate the performance of our present methodologies on further datasets, particularly published by others. Additionally, our promising results motivate investigating more and different CNN architectures, possibly as well many-layered architectures, despite a likely even greater impact on computation times. Finally, since saturation seems to be a good indicator for smoke, it is worthwhile to investigate histogram equalization methods for automatically determining good naive classification thresholds or finding alternative combinations for training CNN models.

**Acknowledgements.** This work was 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. Linux mint 17.3 “rosa” - cinnamon (64-bit) (2006). <https://linuxmint.com/edition.php?id=204>. Accessed 28 Mar 2017
2. Lightning memory-mapped database (2016). <https://symas.com/offerings/lightning-memory-mapped-database>. Accessed 28 Mar 2017

3. OpenCV library (2017). <http://opencv.org/>
4. Python programming language (2017). <https://www.python.org/>
5. Al Sahaf, O.S., Vega-Carrascal, I., Cunningham, F.O., McGrath, J.P., Bloomfield, F.J.: Chemical composition of smoke produced by high-frequency electrosurgery. *Irish J. Med. Sci.* **176**(3), 229–232 (2007)
6. Anthimopoulos, M., Christodoulidis, S., Ebner, L., Christe, A., Mougiakakou, S.: Lung pattern classification for interstitial lung diseases using a deep convolutional neural network. *IEEE Trans. Med. Imaging* **35**(5), 1207–1216 (2016). <http://ieeexplore.ieee.org>
7. Ball, K.: Controlling surgical smoke: A team approach. Information Booklet (2004). <http://www.megadyne.com/pdf/Kay-Ball-Smoke-Booklet.pdf>
8. Calderara, S., Piccinini, P., Cucchiara, R.: Vision based smoke detection system using image energy and color information. *Mach. Vis. Appl.* **22**(4), 705–719 (2011). <http://link.springer.com/10.1007/s00138-010-0272-1>
9. Chen-Rui Chou, M.C.L.: System and Method for Smoke Detection During Anatomical Surgery (2016). <https://www.google.com/patents/US20160239967>
10. Choi, S.H., Kwon, T.G., Chung, S.K., Kim, T.H.: Surgical smoke may be a biohazard to surgeons performing laparoscopic surgery. *Surg. Endosc. Interv. Tech.* **28**(8), 2374–2380 (2014)
11. Chunyu, Y., Jun, F., Jinjun, W., Yongming, Z.: Video fire smoke detection using motion and color features. *Fire Technol.* **46**(3), 651–663 (2010). <http://link.springer.com/10.1007/s10694-009-0110-z>
12. Cosmescu, I.: Automatic smoke evacuator system for a surgical laser apparatus and method therefor (1991). <https://www.google.com/patents/US5199944>
13. Cosmescu, I.: Automatic smoke evacuator and insufflation system for surgical procedures (2006). <https://www.google.com/patents/US20070249990>
14. Dobrogowski, M., Wesolowski, W., Kucharska, M., Sapota, A., Pomorski, L.: Chemical composition of surgical smoke formed in the abdominal cavity during laparoscopic cholecystectomy—assessment of the risk to the patient. *Int. J. Occup. Med. Environ. Health* **27**(2), 314–325 (2014). <http://ijomh.eu/Chemical-composition-of-surgical-smoke-formed-in-the-abdominal-cavity-during-laparoscopic-cholecystectomy-assessment-of-the-risk-to-the-patient,2054,0,2.html>
15. Ferrari, R.J., Zhang, H., Kube, C.R.: Real-time detection of steam in video images. *Pattern Recogn.* **40**(3), 1148–1159 (2007)
16. Gubbi, J., Marusic, S., Palaniswami, M.: Smoke detection in video using wavelets and support vector machines. *Fire Saf. J.* **44**(8), 1110–1115 (2009)
17. Häfner, M., Gangl, A., Liedlgruber, M., Uhl, A., Vécsei, A., Wrba, F.: Combining Gaussian Markov random fields with the discretewavelet transform for endoscopic image classification. In: *Proceedings of the DSP 2009: 16th International Conference on Digital Signal Processing* (2009)
18. Hafner, M., Gangl, A., Liedlgruber, M., Uhl, A., Vecsei, A., Wrba, F.: Endoscopic image classification using edge-based features. In: *2010 20th International Conference on Pattern Recognition*, pp. 2724–2727. IEEE, August 2010. <http://ieeexplore.ieee.org/document/5597011/>
19. Häfner, M., Liedlgruber, M., Uhl, A., Vécsei, A., Wrba, F.: Color treatment in endoscopic image classification using multi-scale local color vector patterns. *Med. Image Anal.* **16**(1), 75–86 (2012). <http://www.sciencedirect.com/science/article/pii/S1361841511000569>
20. He, K., Zhang, X., Ren, S., Sun, J.: Deep Residual Learning for Image Recognition, December 2015. <http://arxiv.org/abs/1512.03385>



21. Hensman, C., Baty, D., Willis, R., Cuschieri, A.: Chemical composition of smoke produced by high-frequency electrosurgery in a closed gaseous environment. *Surg. Endosc.* **12**, 1017 (1998). <http://www.springerlink.com/index/3PDVCC89D248BJT0.pdf>
22. Jia, Y., Shelhamer, E., Donahue, J., Karayev, S., Long, J., Girshick, R., Guadarrama, S., Darrell, T.: Caffe: convolutional architecture for fast feature embedding. In: Proceedings of the 22nd ACM international conference on Multimedia, pp. 675–678. ACM (2014)
23. Kingma, D., Ba, J.: Adam: a method for stochastic optimization. arXiv preprint [arXiv:1412.6980](https://arxiv.org/abs/1412.6980) (2014)
24. Kolesov, I., Karasev, P., Tannenbaum, A., Haber, E.: Fire and smoke detection in video with optimal mass transport based optical flow and neural networks. In: 2010 IEEE International Conference on Image Processing, pp. 761–764. IEEE, September 2010. <http://ieeexplore.ieee.org/document/5652119/>
25. Krizhevsky, A., Sutskever, I., Hinton, G.E.: ImageNet Classification with Deep Convolutional Neural Networks, pp. 1097–1105. Curran Associates Inc., Nevada (2012). <http://papers.nips.cc/paper/4824-imagenet-classification-with-deep-convolutional-neural-networks.pdf>
26. Krizhevsky, A., Sutskever, I., Hinton, G.E.: ImageNet classification with deep convolutional neural networks. In: Pereira, F., Burges, C.J.C., Bottou, L., Weinberger, K.Q. (eds.) *Advances in Neural Information Processing Systems*, vol. 25, pp. 1097–1105. Curran Associates, Inc., Nevada (2012). <http://papers.nips.cc/paper/4824-imagenet-classification-with-deep-convolutional-neural-networks.pdf>
27. Kudo, S., Hirota, S., Nakajima, T., Hosobe, S., Kusaka, H., Kobayashi, T., Himori, M., Yagyuu, A.: Colorectal tumours and pit pattern. *J. Clin. Pathol.* **47**(10), 880–885 (1994). <http://www.ncbi.nlm.nih.gov/pubmed/7962600>, <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC502170>
28. LeCun, Y., Bottou, L., Bengio, Y., Haffner, P.: Gradient-based learning applied to document recognition. *Proc. IEEE* **86**(11), 2278–2324 (1998)
29. Lee, C.Y., Lin, C.T., Hong, C.T., Su, M.T.: Smoke detection using spatial and temporal analyses. *Int. J. Innov. Comput. Inf. Control* **8**(7A), 4749–4770 (2012)
30. Li, Q., Cai, W., Wang, X., Zhou, Y., Feng, D.D., Chen, M.: Medical image classification with convolutional neural network. In: 2014 13th International Conference on Control Automation Robotics & Vision (ICARCV), pp. 844–848. IEEE, December 2014. <http://ieeexplore.ieee.org/document/7064414/>
31. Liedlgruber, M., Uhl, A.: Endoscopic image processing - an overview. In: 2009 Proceedings of 6th International Symposium on Image and Signal Processing and Analysis, pp. 707–712. IEEE, September 2009. <http://ieeexplore.ieee.org/document/5297635/>
32. Buffalo Filter LLC: Surgical Smoke: Education and Training (2017). [http://www.buffalofilter.com/files/7914/1443/3525/Website\\_Training\\_Education\\_Section\\_10\\_27\\_2014.pdf](http://www.buffalofilter.com/files/7914/1443/3525/Website_Training_Education_Section_10_27_2014.pdf)
33. Loukas, C., Georgiou, E.: Smoke detection in endoscopic surgery videos: a first step towards retrieval of semantic events: smoke detection in endoscopic surgery videos. *Int. J. Med. Robot. Comput. Assist. Surg.* **11**(1), 80–94 (2015). <http://doi.wiley.com/10.1002/rcs.1578>

34. Mattes, D., Silajdzic, E., Mayer, M., Horn, M., Scheidbach, D., Wackernagel, W., Langmann, G., Wedrich, A.: Surgical smoke management for minimally invasive (micro)endoscopy: an experimental study. *Surg. Endosc. Interv. Tech.* **24**(10), 2492–2501 (2010)
35. Menes, T., Spivak, H.: Laparoscopy: searching for the proper insufflation gas. *Surg. Endosc.* **14**(11), 1050–1056 (2000). <http://www.ncbi.nlm.nih.gov/pubmed/11116418>
36. Ojo, J., Oladosu, J.: Video-based smoke detection algorithms: a chronological survey. *Comput. Eng. Intell. Syst.* **5**(7), 38–50 (2014)
37. Ott, D.: Smoke production and smoke reduction in endoscopic surgery: preliminary report. *Endosc. Surg. Allied Technol.* **1**(4), 230–232 (1993). <http://www.ncbi.nlm.nih.gov/pubmed/8050026>
38. Park, S.Y., Sargent, D.: Colonoscopic polyp detection using convolutional neural networks. In: International Society for Optics and Photonics, p. 978528, March 2016. <http://proceedings.spiedigitallibrary.org/proceeding.aspx?doi=10.1117/12.2217148>
39. Petscharnig, S., Schöffmann, K.: Deep learning for shot classification in gynecologic surgery videos. In: Amsaleg, L., Gu mundsson, G. ., Gurrin, C., Jónsson, B. ., Satoh, S. (eds.) MMM 2017. LNCS, vol. 10132, pp. 702–713. Springer, Cham (2017). doi:10.1007/978-3-319-51811-4\_57
40. Sahiner, B., Chan, H.-P., Petrick, N., Wei, D., Helvie, M., Adler, D., Goodsitt, M.: Classification of mass and normal breast tissue: a convolution neural network classifier with spatial domain and texture images. *IEEE Trans. Med. Imaging* **15**(5), 598–610 (1996). <http://ieeexplore.ieee.org/document/538937/>
41. Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., Anguelov, D., Erhan, D., Vanhoucke, V., Rabinovich, A.: Going deeper with convolutions. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 1–9 (2015)
42. Takahashi, H., Yamasaki, M., Hirota, M., Miyazaki, Y., Moon, J.H., Souma, Y., Mori, M., Doki, Y., Nakajima, K.: Automatic smoke evacuation in laparoscopic surgery: a simplified method for objective evaluation. *Surg. Endosc.* **27**(8), 2980–2987 (2013). <http://link.springer.com/10.1007/s00464-013-2821-y>
43. Thiébaud, H.P., Knize, M.G., Kuzmicky, P.A., Hsieh, D.P., Felton, J.S.: Airborne mutagens produced by frying beef, pork and a soy-based food. *Food Chem. Toxicol.* **33**(10), 821–828 (1995)
44. Tian, H., Li, W., Wang, L., Ogunbona, P.: A novel video-based smoke detection method using image separation. In: Proceedings - IEEE International Conference on Multimedia and Expo, pp. 532–537 (2012)
45. Toreyin, B.U., Dedeoglu, Y., Cetin, A.E.: Contour Based Smoke Detection in Video Using Wavelets, pp. 1–5. IEEE (2006)
46. Tsui, C., Klein, R., Garabrant, M.: Minimally invasive surgery: national trends in adoption and future directions for hospital strategy. *Surg. Endosc.* **27**(7), 2253–2257 (2013). <http://link.springer.com/10.1007/s00464-013-2973-9>
47. Wu, S., Yuan, F., Yang, Y., Fang, Z., Fang, Y.: Real-time image smoke detection using staircase searching-based dual threshold AdaBoost and dynamic analysis. *IET Image Process.* **9**(10), 849–856 (2015). <http://digital-library.theiet.org/content/journals/10.1049/iet-ipr.2014.1032>

48. Yan, Z., Zhan, Y., Peng, Z., Liao, S., Shinagawa, Y., Zhang, S., Metaxas, D.N., Zhou, X.S.: Multi-Instance deep learning: discover discriminative local anatomies for bodypart recognition. *IEEE Trans. Med. Imaging* **35**(5), 1332–1343 (2016). <http://ieeexplore.ieee.org/document/7398101/>
49. Yuan, F.: Video-based smoke detection with histogram sequence of LBP and LBPV pyramids. *Fire Saf. J.* **46**(3), 132–139 (2011)