Computers in Biology and Medicine 111 (2019) 103351

Contents lists available at ScienceDirect

Computers in Biology and Medicine

journal homepage: www.elsevier.com/locate/compbiomed

MAPGI: Accurate identification of anatomical landmarks and diseased tissue in gastrointestinal tract using deep learning

Timothy Cogan[*], Maribeth Cogan, Lakshman Tamil

Department of Electrical and Computer Engineering, Quality of Life Technology Laboratory, University of Texas at Dallas, Richardson, TX, 75080, USA

ARTICLE INFO

Keywords:
Endoscopy
Gastrointestinal tract
Image enhancement
Machine learning
Neural network
Deep learning

ABSTRACT

Automatic detection of anatomical landmarks and diseases in medical images is a challenging task which could greatly aid medical diagnosis and reduce the cost and time of investigational procedures. Also, two particular challenges of digital image processing in medical applications are the sparsity of annotated medical images and the lack of uniformity across images and image classes. This paper presents methodologies for maximizing classification accuracy on a small medical image dataset, the Kvasir dataset, by performing robust image preprocessing and applying state-of-the-art deep learning. Images are classified as being or involving an anatomical landmark (pylorus, z-line, cecum), a diseased state (esophagitis, ulcerative colitis, polyps), or a medical procedure (dyed lifted polyps, dyed resection margins). A framework for modular and automatic preprocessing of gastrointestinal tract images (MAPGI) is proposed, which applies edge removal, contrast enhancement, filtering, color mapping and scaling to each image in the dataset. Gamma correction values are automatically calculated for individual images such that the mean pixel value for each image is normalized to 90 ± 1 in a 0–255 pixel value range. Three state-of-the-art neural networks architectures, Inception-ResNet-v2, Inception-v4, and NASNet, are trained on the Kvasir dataset, and their classification performance is juxtaposed on validation data. In each case, 85% of the images from the Kvasir dataset are used for training, while the other 15% are reserved for validation. The resulting accuracies achieved using Inception-v4, Inception-ResNet-v2, and NASNet were 0.9845, 0.9848, and 0.9735, respectively. In addition, Inception-v4 achieved an average of 0.938 precision, 0.939 recall, 0.991 specificity, 0.938 F1 score, and 0.929 Matthews correlation coefficient (MCC). Bootstrapping provided NASNet, the worst performing model, a lower bound of 0.9723 accuracy on the 95% confidence interval.

1. Introduction

GASTROINTESTINAL tract related diseases have both high prevalence and high treatment cost. In 2009, 6.9 million upper, 11.5 million lower, and 228,000 biliary endoscopies were performed [1]. One of the jobs of gastroenterologists is to analyze images or video taken along the gastrointestinal (GI) tract. Machine learning algorithms could significantly lower the cost of investigational procedures by providing cognitive enhancement for gastroenterologists and thereby enhancing the speed and quality of image analysis, landmark recognition, and disease detection.

One of the key technologies used to perform digital image analysis is neural networks, in particular, convolutional neural networks (CNN). CNNs are suited for the task of image recognition partly because they are modeled after the mammalian visual cortex [2]. A CNN typically consists of an input layer, an output layer, and at least one hidden convolutional layer. A convolutional layer looks at patches of an image at a time (rather than the whole image at once), and applies the same filters to each patch in order to recognize features of an image such as curves, lines, and edges. Using shared features reduces computation and risk of overfitting [3].

As a result of MediaEval's 2017 Multimedia for Medicine Task, there are several recent papers which present state-of-the-art convolutional networks for gastrointestinal analysis. For this 2017 challenge, researchers designed systems capable of classifying images from the Kvasir dataset in a fast and reliable manner. For example, Agrawal et al. used a support vector machine (SVM) classifier to distinguish images based on their features, where the image features included were baseline features provided with the dataset combined with Agrawal's own neural network extracted features. Two networks pre-trained on ImageNet images were used to extract features – Inception-v3, and VGGNet. It was found that the combination of features extracted by

[*] Corresponding author.
E-mail addresses: timothy.cogan@utdallas.edu (T. Cogan), mxr127730@utdallas.edu (M. Cogan), laxman@utdallas.edu (L. Tamil).

https://doi.org/10.1016/j.compbiomed.2019.103351
Received 19 April 2019; Received in revised form 11 June 2019; Accepted 6 July 2019

both of these networks along with the baseline features provided the best feature set for optimizing the SVM classification accuracy, achieving 96.1% [4]. Pogorelov et al. proposed a similar system but with 2048 ResNet50 features input to a logistic model tree (LMT) classifier, achieving 95.7% accuracy. Pogorelov actually attempted 17 different techniques, but achieved the best accuracy using the pre-trained ResNet50 with LMT classifier [5]. Pogorelov's other techniques explored the efficacy of random forest classifiers, random tree classifiers, Inception-v3 features, and more. Unlike Agrawal and Pogorelov, Petscharnig et al. proposed a purely CNN based classifier inspired by GoogLeNet and achieved a 93.9% accuracy. A remarkable aspect of this CNN is that the authors claim acceptable results are achievable with as few as 400 images. There were also a couple of solutions not based on CNNs which achieved competent results. Liu et al. extracted image features with a bidirectional marginal Fisher analysis (BMFA) and then fed these features into a support vector machine. Naqvi et al. created a separate logistic regression model for each image feature (where the features used were the provided baseline features, plus local binary patterns and Haralick texture features), then used the ensemble method to combine the predictions made by each model and create a final prediction, achieving 94.2% accuracy [5–8].

The goal of this study is to maximize the performance of state-of-the-art CNNs in classifying images from eight classes included in the Kvasir dataset as either anatomical landmarks, diseased states, or medical procedures [9]. Emphasis is given to data preprocessing and augmentation, in order to achieve maximal accuracy without overfitting. We should emphasize that our study is limited since there are only eight classes within the Kvasir dataset, and therefore our system can handle only eight gastrointestinal image classes.

2. Methodology

2.1. Kvasir dataset

The Kvasir dataset is a collection of gastrointestinal tract images taken with an endoscope, and annotated and verified by certified endoscopists. This dataset became available fall of 2017 through the Medical Multimedia Challenge offered by MediaEval, a benchmarking initiative that gives tasks to the research community [9]. Approaches and results of research groups who participated in the MediaEval Medico Challenge can be found in Table 1. Note that the networks proposed by Cogan et al. listed in Table 1 were developed after the challenge was held.

The images in the Kvasir dataset are comprised of eight categories, with 1000 images in each category: three representing anatomical landmarks, three representing pathological states, and two related to lesion-removal. The three anatomical landmark classes are pylorus, z-line, and cecum. The pylorus is the area surrounding the opening between the stomach and the first part of the small intestine. This opening controls the movement of food from the stomach via constricting muscles. Both sides of the pylorus must be examined for a complete gastroscopy. The z-line is the place of transition between the esophagus and the stomach, and is an important landmark to examined for disease detection, as esophagitis typically becomes evident at this location. The cecum (near the ileocecal valve) is the beginning part of the large bowel, and when this landmark is reached, a colonoscopy is considered complete (see Fig. 1). The three diseased states are esophagitis, ulcerative colitis, and polyps. Esophagitis occurs when the esophagus is inflamed, and it results in a break in the mucosa present at the z-line. Ulcerative colitis is a disease which inflames the large bowel. Polyps are outgrowths within the large bowel that can be precancerous. Each of these categories is something key that a gastroenterologist looks for while performing an endoscopy. The two categories related to lesion removal are dyed lifted polyps and dyed resection margins. During polyp removal, dye is used to make the polyp more visible and lifting refers to a technique used to separate polyps from adjacent tissue. Dyed lifted polyp images are taken just prior to polyp removal whereas dyed resection margins are taken after polyp removal [9].

Throughout the dataset, image resolution varies from 720 × 576 pixels to 1920 × 1072 pixels. On some of the images, approximately the leftmost quarter of the image is devoted to annotations and image information, while the anatomical view spans the remaining three quarters of the image. Some images also have a green box in the lower left corner which indicates the location of the endoscope in the GI tract. The images vary in capture angle, resolution, brightness, zoom, and centerpoint. For the task of deep learning, the images are of fairly high complexity, moderate quality, and fairly low volume [9].

2.2. Image preprocessing framework

While robust deep network architectures are becoming ubiquitous with the release of publicly available architectures by technology leaders such as Google, data cleaning and preparation remain as challenges which must be properly addressed in every domain. Consequently, the authors of this paper have developed a framework for preprocessing medical images that is intended to apply specifically to gastrointestinal tract images. The modular adaptive preprocessing for gastrointestinal tract images (MAPGI) framework consists of five primary functional modules: edge removal, contrast enhancement, filtering, color mapping, and scaling which are depicted in Fig. 2. An image which has been processed by MAPGI is shown in Fig. 3. Within the framework, color images are represented in *YUV* color space because the *Y* component solely encodes image luminance. In *RGB* color space, image luminance is described by each of the three color channels. For *YUV* color space, the *Y* component can be treated similar to a greyscale image, so the MAPGI modules produce similar effects whether or not the input image contains color [10].

Medical images commonly have undesirable edge artifacts or annotations that are distractions to the learning task at hand. The MAPGI edge removal module is capable of handling numerous types of edge artifacts, such as black borders, white streaks caused by scanning instruments, and annotations. In the case of the Kvasir dataset, MAPGI was used to remove all black borders and areas devoted to annotations (no physiological information present). Black border removal was

Table 1
Applications of machine learning in GI tract analysis research using the Kvasir dataset.

Author	Year	Technique	Portion Analyzed	ACC	PREC	REC	SPEC	F1	MCC
This study	2019	Inception-v4	Cropped/resized to 299 × 299	0.9845	0.938	0.939	0.991	0.938	0.929
This study	2019	Inception-ResNet-v2*	Cropped/resized to 299 × 299	0.9848	0.940	0.939	0.991	0.939	0.930
This study	2019	NASNet*	Cropped/resized to 299 × 299	0.9735	0.893	0.898	0.985	0.892	0.879
Agrawal et al. [4]	2017	Base feat. + Incep-v3 + VGGNet + SVM	Whole img, res. 244 × 244	0.961	0.847	0.852	0.978	0.847	0.827
Pogorelov et al. [5]	2017	ResNet50 feat. + LMT	Img feat., res. N/A	0.957	0.829	0.826	0.975	0.826	0.802
Naqvi et al. [6]	2017	Base feat. + LBP + Haralick + LR	Img feat., whole img	0.942	0.767	0.774	0.966	0.767	0.736
Petscharnig et al. [7]	2017	GoogLeNet-based CNN	128 × 128 img patches, 7/img	0.939	0.755	0.755	0.965	0.755	0.720
Liu et al. [8]	2017	BMFA + SVM	Whole img, BMFA feat.	0.926	0.703	0.703	0.958	0.703	0.660

*Inception-ResNet-v2 and NASNet were only evaluated on a single set of 85/15 splits.

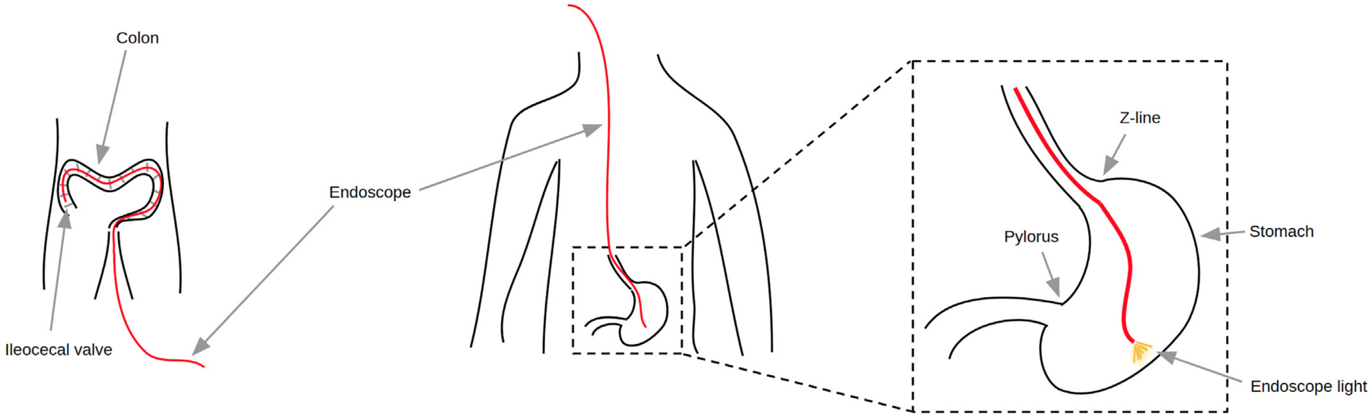

Fig. 1. Image depicting a colonoscopy on the left and an upper endoscopy procedure on the right. During either procedure a biopsy sample may be taken to detect malignant tissue or harmful bacteria.

performed by deleting any row or column from the image where all pixel values were below a set threshold. Areas devoted to annotation (white text on a black background) were removed by deleting columns containing fewer than a set number of pixels with intensity above a defined threshold. That is, any row or column where the majority of pixels are black and a minority are white are removed, because these represent annotations. Although relatively simple, these techniques provided a fast and effective means to deal with undesired artifacts within the Kvasir dataset.

The MAPGI contrast enhancement module consists of two key tools used to correct an image's brightness and contrast: Contrast-Limited Adaptive Histogram Equalization (CLAHE), and Mean-Approximated Gamma Value Adjustment (MAGVA). CLAHE is a relatively popular technique for contrast enhancement and can be easily applied via a publicly available function from OpenCV [11]. CLAHE works by first dividing an image into *NxN* tiles (e.g. sections of 8 × 8 pixels). Next, CLAHE enhances image contrast by performing histogram equalization on each tile. Histogram equalization is a process by which the pixel values of a given tile are shifted so that the pixel histogram has a uniform distribution. Prior to calculating a histogram equalization transform, however, contrast limiting is performed to clip high pixel histogram frequencies and thereby reduce noise amplification [12]. MAGVA is a function developed by the authors of this paper which compares the mean pixel value of the current image to a desired mean pixel value, estimates a gamma value needed to correct the current image's brightness, and then applies a gamma correction. The function performs this process recursively until the image's mean pixel value is within a specified range of the desired mean. One or both of these tools (MAGVA or CLAHE) can be used to improve image contrast, and the use of either will depend on the preprocessing needs of the dataset in question. Because it operates on tiles of pixels, subsets of an overall image, CLAHE is effective at improving contrast within small regions of an image. CLAHE is ideal in an image that has appropriate contrast for all but a few patches of pixels. Alternatively, MAGVA enhances contrast across an entire image in a way that depends only on pixel value and not pixel location. MAGVA is a global enhancement technique which is ideal when an entire image is too dark or too light. For the case of the Kvasir dataset, MAGVA was used to create uniformity in image brightness and contrast across the dataset, setting the pixel mean value of each image to 90 ± 1.

A derivation for MAGVA is briefly described here. Suppose the desired output pixel mean for an image is b. That is, we desire:

$$\frac{1}{n}\sum_{i=1}^{n} O_i = b \tag{1}$$

O_i represents the ith pixel value following a gamma correction. Now inserting the gamma correction equation [13] gives us:

$$\frac{255}{n}\sum_{i=1}^{n}\left(\frac{I_i}{255}\right)^{\lambda} = b \tag{2}$$

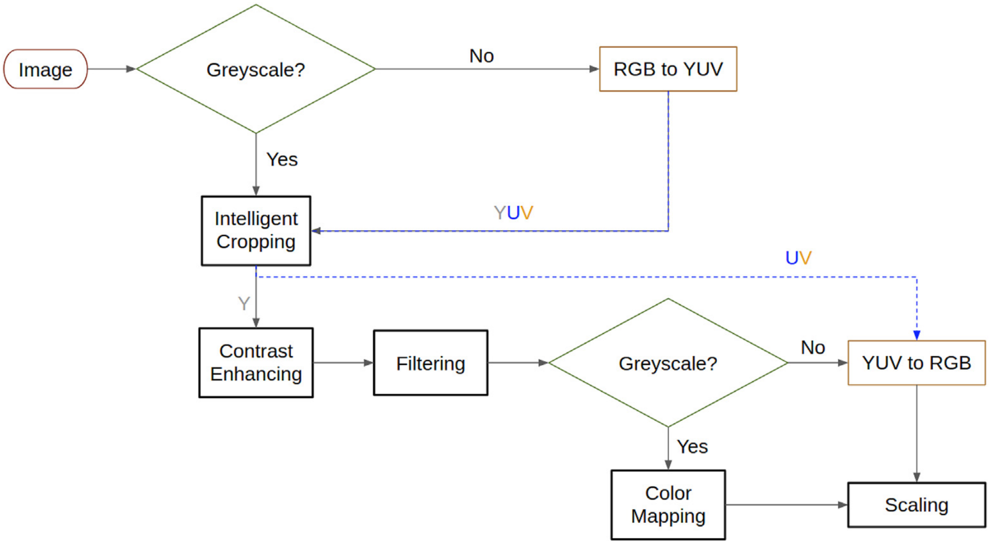

Fig. 2. Flow diagram depicting image preprocessing modules within MAPGI.

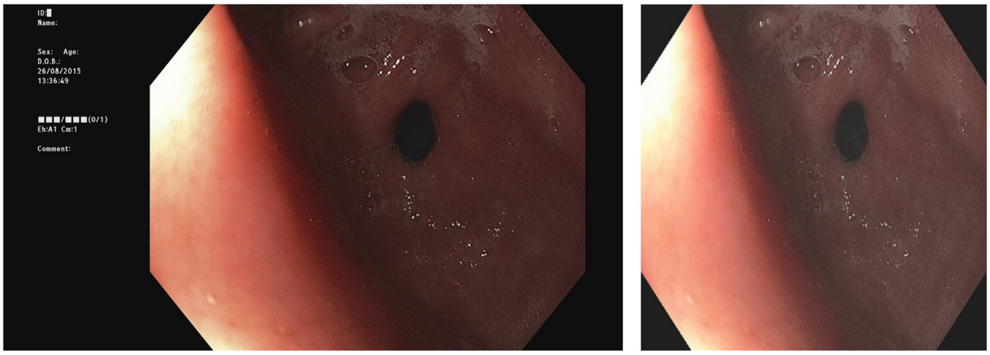

Fig. 3. Left: Original image of Pylorus from the Kvasir dataset, with annotations included. Right: Cropped/adjusted image, after application of MAPGI.

I_i represents the ith pixel value prior to gamma correction. To approximate the solution for the desired gamma correction coefficient λ, the input pixels I_i are replaced by the average pixel value of the image prior to correction, a, giving:

$$\frac{255}{n}\sum_{i=1}^{n}\left(\frac{a}{255}\right)^{\lambda} = b \tag{3}$$

$$\left(\frac{a}{255}\right)^{\lambda} = \frac{b}{255} \tag{4}$$

$$\lambda \ln\frac{a}{255} = \ln\frac{b}{255} \tag{5}$$

$$\lambda = \frac{\ln b - \ln 255}{\ln a - \ln 255} \tag{6}$$

Solving for the approximated lambda via equation (6), using that lambda value to apply gamma correction, and then repeating this process until the image mean a is sufficiently close to b constitutes MAGVA. The procedure for applying MAGVA can be seen in Algorithm 1, and an example application can be see in Fig. 4.

Algorithm 1. Mean-Approximated Gamma Value Adjustment (MAGVA)

```
mean = get_image_mean(image)
while mean<lower_bound or mean>upper_bound do
    gamma = ln(desired_mean / 255) / ln(mean / 255)
    image = adjust_gamma(image, gamma)
    mean = get_image_mean(image)
```

Following contrast enhancement, MAPGI filtering can utilize low-pass, high-pass, or band-pass filtering in order to reduce out-of-band noise. In the case of a *YUV* color image, the filtering is only applied to the *Y* component. For the images from the Kvasir dataset, a relatively simple low-pass filter was applied by convolving {[0.1, 0.1, 0.1],[0.1, 1, 0.1],[0.1, 0.1, 0.1]}/1.8 across the entire image. Each set of square brackets *[]* indicates a unique row of the kernel and *1.8* is simply a normalization factor. In principle, Butterworth, Gaussian, and other 2D filter types may be used [14].

Following filtering, MAPGI provides a color mapping module which is useful for enhancing potential greyscale images. In contrast with greyscale images, color mapped images can convey enhanced visual dynamic range by leveraging the entire RGB colorspace. In addition, converting 12-bit or 16-bit greyscale images into 8-bit greyscale (24-bit RGB) images will result in a loss of visual information, but converting 12-bit greyscale to 24-bit color-mapped RGB will preserve some of the otherwise lost visual information [15]. A 12-bit greyscale to 24-bit RGB conversion may be convenient prior to feeding images into a pre-trained convolutional neural network [16]. In the case of the Kvasir dataset, no color mapping was applied since all of the Kvasir images were already 24-bit color images.

MAPGI scaling utilizes bilinear interpolation. Bilinear interpolation is a relatively straightforward technique for two dimensional interpolation whereby linear interpolation is executed along each axis. Interpolation is first performed along one axis, and then interpolation is performed along the remaining axis by use of the values produced via interpolation along the first axis [17]. For the Kvasir images, resizing was necessary because Inception-v4, Inception-ResNet-v2, and NASNet all expect 299 × 299 images.

Lastly, following MAPGI, dataset augmentation was performed to address the small number of samples available for training the network and reduce overfitting. This was done by randomly mirroring images along both the horizontal and vertical axes. In effect, this helps the network learn what anatomical structures look like when viewed from

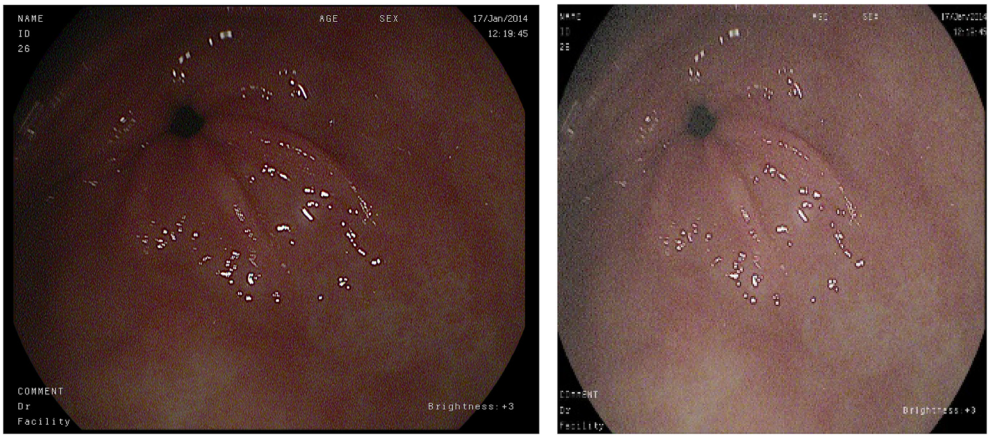

Fig. 4. Left: Original image of Pylorus from the Kvasir dataset with very low mean pixel value. Right: Contrast enhanced image after application of MAPGI.

Table 2
Neural network hyperparameters.

Optimizer	Decay	Momentum	Epsilon	Learning rate	Decay type	Decay factor	Epochs per decay	End rate	Weight decay
RMSprop	0.9	0.9	1.0	0.01	Exponential	0.94	2	0.0001	0.00004

different perspectives [18].

2.3. Convolutional neural network

Three different models, Inception-v4, Inception-ResNet-v2, and NASNet, were trained on the Kvasir dataset. Inception-ResNet-v2 and Inception-v4 were trained using High Performance Computing (HPC) resources provided by Texas Advanced Computing Center (TACC), while NASNet was trained on resources provided by Google Cloud Platform. Hyperparameters used for training these models are shown in Table 2. Each of these models are state-of-the-art convolutional neural networks which have demonstrated strong performance in image classification tasks. Inception-v4 and Inception-ResNet-v2 were both described by Google engineers in 2016 and offer similar performance. However, Inception-ResNet-v2 appears to offer slightly better performance on ImageNet images due to residual connections [19]. Unlike normal connections, residual connections learn deviations from an identity function. That is, whereas a normal connection will try to optimize *f(x)*, a residual connection will try to optimize *f(x)* + *x*. Empirically, these residual connections may lead to better results than normal connections. Inception-v4, hence the name, was created as an improvement to Inception-v3 with a goal of simplifying the overall Inception architecture while providing improved results. Inception-v4 is also a successor to GoogLeNet, also known as Inception-v1. An overview of the Inception-v4 and Inception-ResNet-v2 architectures can be seen in Fig. 5 [19].

NASNet is a slightly more recent network, described by Zoph et al. (also Google engineers) in 2017. NASNet was created by use of the Neural Architecture Search (NAS) framework which provides an algorithm for finding optimal neural network architectures. In other words, the NASNet architecture is not directly designed by people. NASNet offers higher performance on the ILSVRC 2012 dataset than do Inception-v4 and Inception-ResNet-v2, 96.2% versus 95.2% and 95.3% accuracy, respectively. However, this increase of performance comes at the cost of network size and necessary multiply-accumulate operations. NASNet-A (6 @ 4032), the NASNet variant which achieved 96.2% accuracy, requires 88.9 million parameters and 23.8 billion multiply-accumulates versus Inception-ResNet-v2's 55.8 million parameters and 13.2 billion multiply-accumulates. An overview of the NASNet ImageNet architecture is shown in Fig. 5 [20].

3. Results

Out of 8000 images, 15% or 1,200 were used for validation while 85% or 6,800 were used for training. With data set augmentation, effectively four times 6,800 or 27,200 images were used for training. Accuracies of 0.9845, 0.9848, and 0.9735 were achieved with Inception-v4, Inception-ResNet-v2, and NASNet, respectively. Other performance metrics aside from accuracy can be found in Table 1. Unsurprisingly, Inception-v4 and Inception-ResNet-v2 showed similar performance. We believe that the NASNet model did not perform as well due to the size of the model and a relatively small number of training images. As mentioned previously, whereas Inception-ResNet-v2 contains 55.8 million parameters, NASNet contains 88.9 million parameters, almost 1.6 times the number of parameters as Inception-ResNet-v2. On a limited dataset such as the Kvasir dataset, NASNet is more likely to overfit the training data, resulting in diminished performance on the testing data. With more training data, the NASNet model might perform as good as if not better than the Inception-v4 and Inception-ResNet-v2 models. The performance of these networks is further compared in Fig. 6, where we can see an insignificant performance difference between Inception-v4 and Inception-ResNet-v2.

To further increase confidence in these results, Inception-v4 was trained and evaluated on an additional two random 85/15 splits. Inception-v4 was chosen because it seemed to perform as well or better than the other two models, Inception-ResNet-v2 and NASNet, but is the smallest of the three models. Smaller models are ideal from computational and overfitting points of view. Fig. 7 shows that all three evaluations with Inception-v4 are relatively consistent. The results in Table 1 represent the average of these three evaluations, but the bootstrapping results depicted in Fig. 6 were generated from only the first Inception-v4 evaluation.

The confusion matrix in Table 3 shows the classification results on validation data for the eight classes by Inception-v4 for a single experiment. Predicted classes are on the vertical axis, and true classes on the horizontal axis. Perfect classification accuracy would result in a

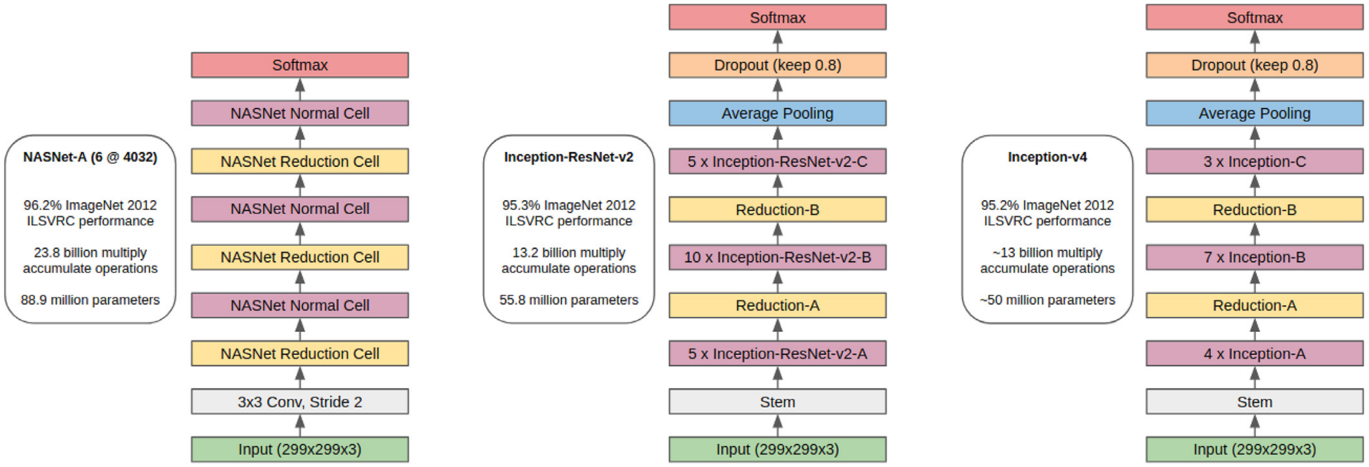

Fig. 5. Architectures for NASNet-ImageNet, Inception-ResNet-v2, and Inception-v4 are depicted from left to right, respectively. The number of characteristic layers (e.g. Inception-A, Inception-ResNet-v2-B, NASNet normal cell) at each step is a hyperparameter that may change depending on the initial training dataset. These architectural diagrams are adapted from the original papers in which the networks were introduced [19,20].

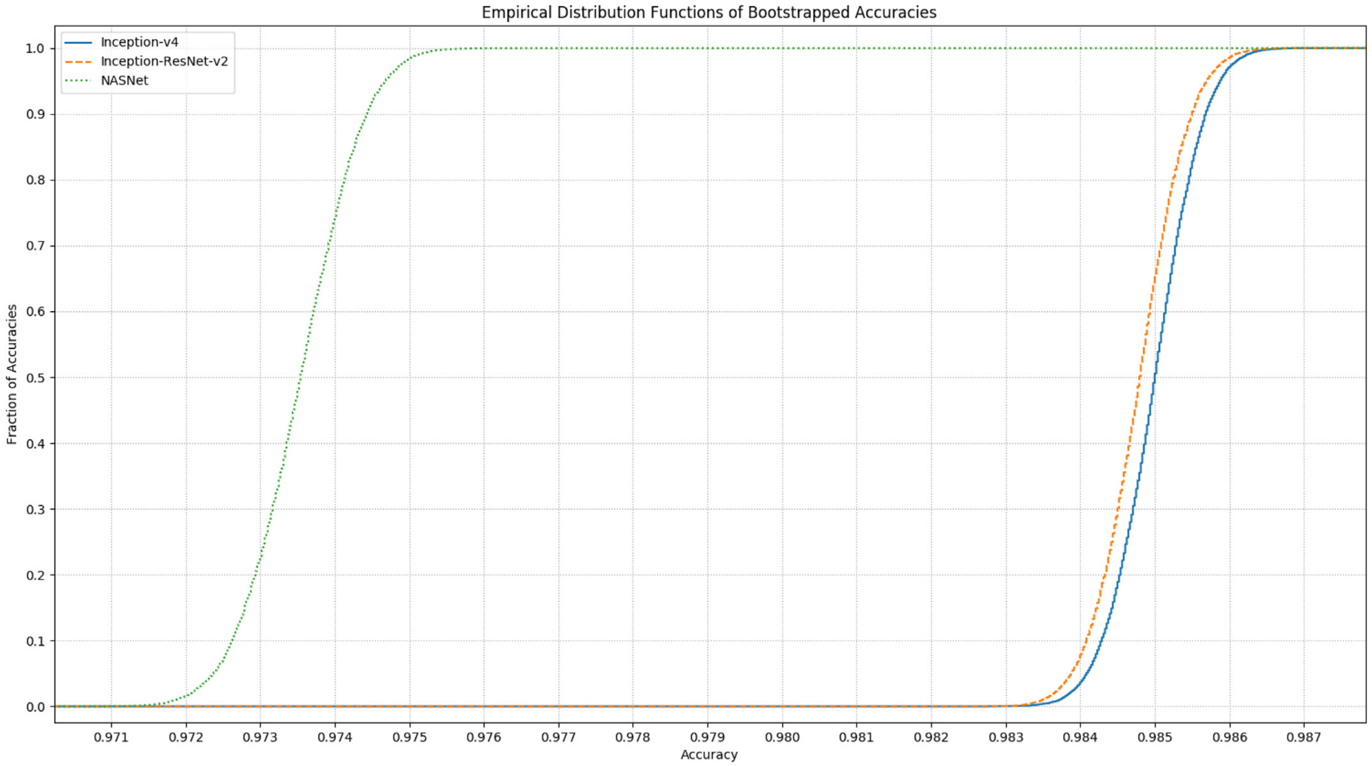

Fig. 6. Bootstrapped accuracies for the 3 different networks. Bootstrapped datasets were generated by sampling the validation dataset with replacement, and accuracies corresponding to each of these generated datasets determine empirical distribution functions. Although Inception-v4 and Inception-ResNet-v2 models outperformed the NASNet model, the performance difference between these 2 models is not statistically significant because there is substantial overlap between their corresponding empirical distribution functions.

diagonal matrix. It is evident from the confusion matrix that the primary classes that the network confuses are esophagitis and z-line. This is understandable because these two classes contain images of the same anatomical location, which in one case are diseased, and the other case healthy [9].

4. Discussion

As shown in Table 1, the results presented in this paper are competitive with the results seen in previous studies. However, these results cannot be directly compared because the training and validation sets used here are not identical to the training and validation sets used by the other authors. A fair comparison would require predefined training and validation data sets. Regardless, the systems presented by this paper appear to perform well on each of the metrics provided in Table 1. Like this study, other studies have also leveraged deep neural networks such as Inception-v3, VGGNet, and ResNet50. Unlike this study, many of the other studies simply used the deep networks for feature extraction and relied on another classifier (e.g. SVM) to assign labels. In addition, the neural networks we explored are generally larger, costlier to train and evaluate, and have been developed more recently. Lastly, we believe our cropping and contrast enhancement preprocessing techniques are unique among existing literature [4–8].

There are many strategies that could be employed to improve upon the work here described. First, better results could be achieved by making use of the extracted image features provided along with the Kvasir dataset. Using these features would augment the feature set extracted by the CNN, thus giving a more complete picture of the information contained within the image. Second, the network could be improved by employing more relevant transfer learning. In transfer learning, a network pre-trained on other image datasets is fine-tuned to the particular set of images and task at hand. The benefit of using a pre-

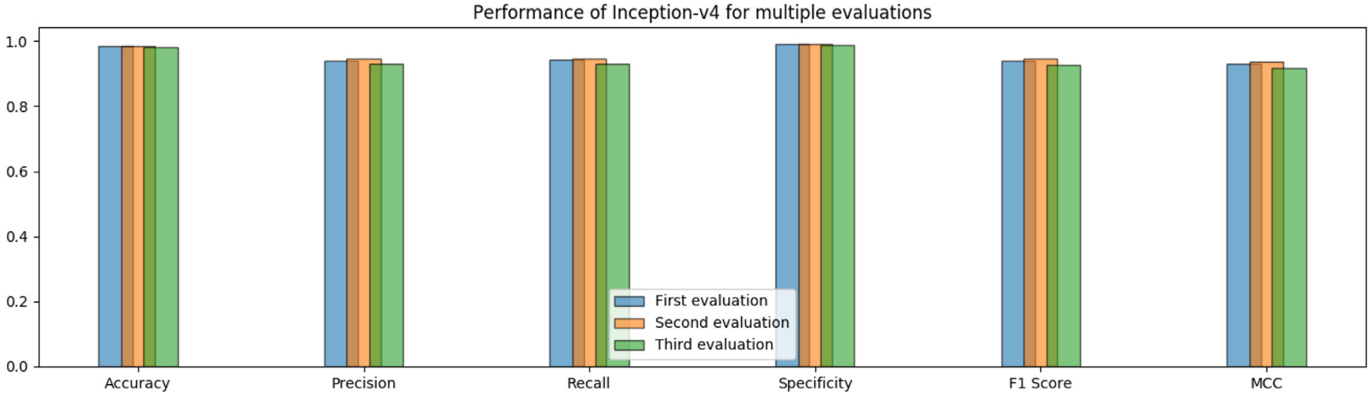

Fig. 7. Performance of Inception-v4 over 3 independent evaluations on random 85/15 data splits, showing that the results are consistent.

Table 3
Confusion matrix for Inception-v4 classification results of 8-classes of data from the Kvasir dataset.

Pred, True	Cecum	D.L.P	D.R.M.	Esoph.	Polyps	Pylorus	U.Colitis	Z-line
Cecum	152	1	0	0	4	0	6	0
D.L.P.	0	132	3	0	2	0	0	0
D.R.M.	0	1	163	0	0	0	0	0
Esoph.	0	0	0	111	0	0	0	11
Polyps	0	1	0	0	141	0	0	0
Pylorus	0	0	0	1	2	139	0	0
U.Colitis	2	0	0	0	2	0	152	0
Z-line	0	0	0	36	0	0	0	138

trained model is that features common to most images (lines, curves, edges, etc.) the network already knows how to identify, and thus, the network has less to learn by the time it sees the particular set of images in question [21]. Although the networks used in this paper were pre-trained on ImageNet images, ideally for the task of identifying anatomical landmarks and diseases, the network would be pre-trained on other medical image datasets where the images have similar features. Third, better results could be achieved through better tuning of the neural network hyperparameters such as dropout rate, learning rate, gradient descent technique, activation function, regularization method, and network architecture [22]. Fourth, further variations of MAPGI could be applied. Different filter types, scaling methods, or contrast enhancement techniques could lead to higher classification accuracy. Finally, additional methods of dataset augmentation could be performed such as cropping subsets of images. The advantage of cropping an image to use only a subset is that the key portions of the image determining its class can be isolated, and used solely in the training of the network [23]. In this way, the network does not waste resources learning portions of an image that are not particularly relevant to the image class. This could especially help in the case of polyp detection. If images were cropped to include only the polyp, the network would be forced to learn the features of the polyp, and would not be distracted by the other anatomical landmarks surrounding the polyp.

5. Conclusion

In spite of the previously mentioned modifications that could be used to improve the network performance, our proposed system has demonstrated performance which is competitive to other state-of-the-art classifiers. With high levels of accuracy and precision we are able to distinguish eight classes of gastrointestinal landmarks and diseases. Systems such as ours could greatly aid medical professionals in diagnosis, thereby reducing the overall cost and time of investigational procedures. Furthermore, the authors believe that the MAPGI framework here described provides a reference by which future preprocessing systems may be built, since operations of cropping and contrast enhancement are important across multiple domains.

In addition, we have a high level of confidence in the results presented here. All three state-of-the-art neural networks, Inception-v4, Inception-ResNet-v2, and NASNet, achieved comparable results under the same image preprocessing pipeline, which is expected since these networks are near-equal performers on the ILSVRC 2012 dataset. Also, bootstrapping the results from these networks indicates a high probability that each one of these networks is a strong performer. Lastly, consistency in the two additional train/test evaluations of Inception-v4 gives us confidence that our system is robust.

Conflicts of interest

None.

Acknowledgment

The authors would like to acknowledge the Texas Advanced Computing Center (TACC) at The University of Texas at Austin for providing HPC resources that have contributed to the research results reported within this paper. URL: http://www.tacc.utexas.edu.

References

[1] A.F. Peery, E.S. Dellon, J. Lund, S.D. Crockett, C.E. McGowan, W.J. Bulsiewicz, L.M. Gangarosa, M.T. Thiny, K. Stizenberg, D.R. Morgan, et al., Burden of gastrointestinal disease in the United States: 2012 update, Gastroenterology 143 (5) (2012) 1179–1187.

[2] S. Lau, A Walkthrough of Convolutional Neural Network - Hyperparameter Tuning, (Jul 2017) [Online]. Available: https://towardsdatascience.com/a-walkthrough-of-convolutional-neural-network-7f474f91d7bd.

[3] M. Abadi, A. Agarwal, P. Barham, E. Brevdo, Z. Chen, C. Citro, G.S. Corrado, A. Davis, J. Dean, M. Devin, S. Ghemawat, I. Goodfellow, A. Harp, G. Irving, M. Isard, Y. Jia, R. Jozefowicz, L. Kaiser, M. Kudlur, J. Levenberg, D. Mané, R. Monga, S. Moore, D. Murray, C. Olah, M. Schuster, J. Shlens, B. Steiner, I. Sutskever, K. Talwar, P. Tucker, V. Vanhoucke, V. Vasudevan, F. Viégas, O. Vinyals, P. Warden, M. Wattenberg, M. Wicke, Y. Yu, X. Zheng, TensorFlow: Large-Scale Machine Learning on Heterogeneous Systems, (2015) software available from: tensorflow.org. [Online]. Available: http://tensorflow.org/.

[4] T. Agrawal, R. Gupta, S. Sahu, and C. E. Wilson, “Scl-umd at the Medico Task-Mediaeval 2017: Transfer Learning Based Classification of Medical Images.”.

[5] K. Pogorelov, K.R. Randel, C. Griwodz, S.L. Eskeland, T. de Lange, D. Johansen, C. Spampinato, D.-T. Dang-Nguyen, M. Lux, P.T. Schmidt, et al., Kvasir: a multi-class image dataset for computer aided gastrointestinal disease detection, Proceedings of the 8th ACM on Multimedia Systems Conference, ACM, 2017, pp. 164–169.

[6] S. S. A. Naqvi, S. Nadeem, M. Zaid, and M. A. Tahir, “Ensemble of Texture Features for Finding Abnormalities in the Gastro-Intestinal Tract.”.

[7] S. Petscharnig, K. Schöffmann, M. Lux, An Inception-like Cnn Architecture for Gi Disease and Anatomical Landmark Classification, (2017).

[8] Y. Liu, Z. Gu, and W. K. Cheung, “Hkbu at Mediaeval 2017 Medico: Medical Multimedia Task.”.

[9] K. Pogorelov, K.R. Randel, C. Griwodz, S.L. Eskeland, T. de Lange, D. Johansen, C. Spampinato, D.-T. Dang-Nguyen, M. Lux, P.T. Schmidt, M. Riegler, P. Halvorsen, Kvasir: a multi-class image dataset for computer aided gastrointestinal disease detection, Proceedings of the 8th ACM on Multimedia Systems Conference, Ser. MMSys'17, ACM, New York, NY, USA, 2017, pp. 164–169 [Online]. Available: http://doi.acm.org/10.1145/3083187.3083212.

[10] L. Torres, J.-Y. Reutter, L. Lorente, The importance of the color information in face recognition, Image Processing, 1999. ICIP 99. Proceedings. 1999 International Conference on, vol. 3, IEEE, 1999, pp. 627–631.

[11] Clahe Class Reference.” [Online]. Available: https://docs.opencv.org/3.1.0/d6/db6/classcv_1_1CLAHE.html.

[12] K. Zuiderveld, Contrast limited adaptive histogram equalization, Graphics Gems IV, Academic Press Professional, Inc., 1994, pp. 474–485.

[13] S.-C. Huang, F.-C. Cheng, Y.-S. Chiu, Efficient contrast enhancement using adaptive gamma correction with weighting distribution, IEEE Trans. Image Process. 22 (3) (2013) 1032–1041.

[14] R.A. Haddad, A.N. Akansu, A class of fast Gaussian binomial filters for speech and image processing, IEEE Trans. Signal Process. 39 (3) (1991) 723–727.

[15] Mpl Colormaps. https://bids.github.io/colormap/. Accessed: 2017-11-11.

[16] T. Cogan, M. Cogan, L. Tamil, Rams: remote and automatic mammogram screening, Comput. Biol. Med. 107 (2019) 18–29.

[17] K.T. Gribbon, D.G. Bailey, A novel approach to real-time bilinear interpolation, Electronic Design, Test and Applications, Proceedings. DELTA 2004. Second IEEE International Workshop on, IEEE, 2004, pp. 126–131.

[18] M. Kim, J. Zuallaert, W. De Neve, Towards novel methods for effective transfer learning and unsupervised deep learning for medical image analysis, Doctoral Consortium (DCBIOSTEC 2017), 2017, pp. 32–39.

[19] C. Szegedy, S. Ioffe, V. Vanhoucke, Inception-v4, inception-resnet and the impact of residual connections on learning, CoRR abs/1602 (2016) 07261[Online]. Available: http://arxiv.org/abs/1602.07261.

[20] B. Zoph, V. Vasudevan, J. Shlens, Q.V. Le, Learning transferable architectures for scalable image recognition, CoRR abs/1707 (2017) 07012[Online]. Available: http://arxiv.org/abs/1707.07012.

[21] H.-W. Ng, V.D. Nguyen, V. Vonikakis, S. Winkler, Deep learning for emotion recognition on small datasets using transfer learning, Proceedings of the 2015 ACM on International Conference on Multimodal Interaction, ACM, 2015, pp. 443–449.

[22] T. Domhan, J.T. Springenberg, F. Hutter, Speeding up automatic hyperparameter optimization of deep neural networks by extrapolation of learning curves, IJCAI, 2015, pp. 3460–3468.

[23] H. Furukawa, Deep Learning for Target Classification from Sar Imagery: Data Augmentation and Translation Invariance, (2017) arXiv preprint arXiv:1708.07920.