Angiodysplasia Detection and Localization Using Deep Convolutional Neural Networks

Alexey Shvets   SHVETS@MIT.EDU

Institute for Medical Engineering and Science
Massachusetts Institute of Technology
Boston, MA 02142, United States of America

Vladimir Iglovikov   IGLOVIKOV@GMAIL.COM

Lyft Inc.
San Francisco, CA 94107, United States of America

Alexander Rakhlin   RAKHLIN@NEUROMATION.IO

Neuromation OU
Tallinn, 10111 Estonia

Alexandr A. Kalinin   AKALININ@UMICH.EDU

Department of Computational Medicine and Bioinformatics
University of Michigan
Ann Arbor, MI 48109, United States of America

Abstract

Accurate detection and localization for angiodysplasia lesions is an important problem in early stage diagnostics of gastrointestinal bleeding and anemia. Gold-standard for angiodysplasia detection and localization is performed using wireless capsule endoscopy. This pill-like device is able to produce thousand of high enough resolution images during one passage through gastrointestinal tract. In this paper we present our winning solution for MICCAI 2017 Endoscopic Vision SubChallenge: Angiodysplasia Detection and Localization its further improvements over the state-of-the-art results using several novel deep neural network architectures. It address the binary segmentation problem, where every pixel in an image is labeled as an angiodysplasia lesions or background. Then, we analyze connected component of each predicted mask. Based on the analysis we developed a classifier that predict angiodysplasia lesions (binary variable) and a detector for their localization (center of a component). In this setting, our approach outperforms other methods in every task subcategory for angiodysplasia detection and localization thereby providing state-of-the-art results for these problems. The source code for our solution is made publicly available at https://github.com/ternaus/angiodysplasia-segmentation

1. Introduction

Angiodysplasia (AD) is the most common vascular lesion of the gastrointestinal (GI) tract in the general population (Foutch et al., 1995). This condition may be asymptomatic, or it may cause gastrointestinal bleeding or and anemia (Regula et al., 2008). Small bowel angiodysplasia may account for 30-40% of cases of GI bleeding of obscure origin (OGIB). In a retrospective colonoscopic analyses study, it was shown that 12.1% of 642 persons without symptoms of irritable bowel syndrome (IBS), and 11.9% of those with IBS had colonic angiodysplasia (Akhtar et al., 2006). In patients older than 50 years, small bowel AD is the most likely reason of OGIB (Sidhu et al., 2008). Liao et al. (2010) performed a systematic review of all original articles relevant to wireless capsule endoscopy (WCE)
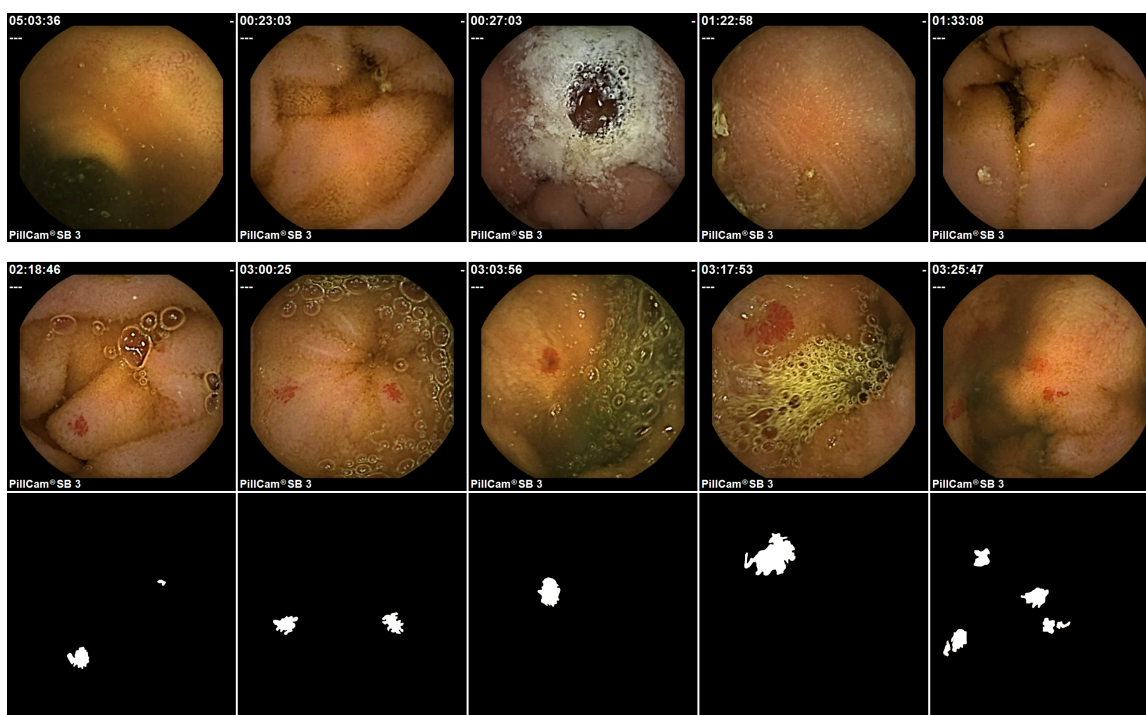
Figure 1: Sample images from the training set for Angiodysplasia detection and localization challenge (MICCAI 2017 Endoscopic Vision Challenge). The upper row corresponds to normal images. In the middle row the images contain angiodysplasia area represented as red spots. The down row contains masks for angiodysplasia from the middle row.

for the evaluation of patients with small bowel signs and symptoms published between 2000 and 2008. A total of 227 studies involving 22 840 procedures were included. OGIB (overt and occult) was the most common indication (66.0%) and AD was the most common cause (50.0%) of bleeding in those patients. In another study, small bowel AD lesions were the most common cause (35%) of severe lifethreatening overt OGIB (Lecleire et al., 2012). Lesions are often multiple, and frequently involve the cecum or ascending colon, although they can occur at other places (Sami et al., 2014).

The diagnosis of a vascular anomaly can be based upon endoscopic findings, histologic characteristics, or association with systemic diseases. Commonly used endoscopic modalities for assessment of the small bowel include WCE, push enteroscopy, deep small bowel enteroscopy [doubleballoon enteroscopy (DBE), singleballoon enteroscopy (SBE) and spiral enteroscopy (SE)] or intraoperative enteroscopy (Sami et al., 2014). Wireless capsule endoscopy (see Fig.2) is the preferred firstline investigation for the small bowel in the context of GI bleeding as it is safe, acceptable and has significantly higher or at least equivalent yield for lesions when compared with other, more invasive modalities like push enteroscopy, mesenteric angiography and intraoperative enteroscopy (Triester et al., 2005; Marmo et al., 2005). Last generation of these pill-like devices can produce more than 60 000 images

with a resolution of approximately 520×520 pixels. However, only 69% of angiodysplasias are detected by gastroenterologist experts during the reading of WCE videos, and blood indicator software (provided by WCE provider like Given Imaging), in the presence of angiodysplasias, presents sensitivity and specificity values of only 41% and 67%, respectively (MICCAI 2017 Endoscopic Vision Challenge). Therefore, there is a compelling need to improve accuracy of AD detection and localization for clinical practice. In this work we apply modern deep learning techniques for automatic detection and localization of angiodysplasia.

There is a number of computer vision-based methods developed for the video capsule endoscopy analysis (Iakovidis and Koulaouzidis, 2015), including rule-based and conventional machine learning algorithms that are applied to extracted color, texture, and other features (Mackiewicz et al., 2008; Karargyris and Bourbakis, 2010; Szczypiski et al., 2014). Recently, deep learning-based approaches demonstrated performance improvements over conventional machine learning methods for many problems in biomedicine (Ching et al., 2018; Kalinin et al., 2018). In the domain of medical imaging, convolutional neural networks (CNN) have been successfully used, for example, for breast cancer histology image analysis (Rakhlin et al., 2018), bone disease prediction (Tiulpin et al., 2018) and age assessment (Iglovikov et al., 2017b), and other problems (Ching et al., 2018). In the analysis of video capsule endoscopy, deep learning has recently demostrated promising results for polyp detection (Tajbakhsh et al., 2015; Yuan and Meng, 2017; Byrne et al., 2017; Murthy et al., 2017).

In this paper, we present a deep learning-based solution for angiodysplasia lesions segmentation from video capsule endoscopy that achieves state-of-the-art results in both binary and multi-class setting. We used this method to produce a submission to the MICCAI 2017 Endoscopic Vision SubChallenge: Angiodysplasia detection and localization (MICCAI 2017 Endoscopic Vision Challenge) that placed first, winning the competition. Here we describe the details of that solution based on a modification of the U-Net model (Ronneberger et al., 2015; Iglovikov et al., 2017a). Moreover, we provide further improvements over this solution utilizing recent deep architectures: TernausNet (Iglovikov and Shvets, 2018) and AlbuNet (Shvets et al., 2018). To our knowledge this is first paper that try to examine angiodysplasia detection and localization using deep learning tools as a results it will serve as state of the art for other investigators.

Figure 2: A wireless capsule endoscope (WCE): (1) Optical Dome, (2) Lens holder,(3) Lens, (4) Illuminating LEDs, (5) CMOS imager, (6) battery, (7) ASIC RF transmitter, (8) Antenna. Source: http://www.yalemedicalgroup.org/news/ymg proctor.html

2. Methods

2.1 Dataset description and preprocessing

A wireless capsule endoscope, is a disposable plastic capsule that weights 3.7g and measures 11mm in diameter and 26mm in length, Fig.2. Image features include a 140 degree field of view, 1:8 magnification, 1 to 30mm depth of view, and a minimum size of detection of about 0.1mm. The capsule is passively propelled through the intesine by peristalsis while transmitting color images. Last generation of this device is able to acquire more than 60,000 images with a resolution of approximately 520×520 pixels (Mishkin et al., 2006).

The dataset consists of 1200 color images obtained with WCE, Fig.2. The images are in 24-bit PNG format, with 576×576 pixel resolution. The dataset is split into two equal parts, 600 images for training and 600 for evaluation. Each subset is composed of 300 images with apparent AD and 300 without any pathology. The training subset is annotated by human expert and contains 300 binary masks in JPEG format of the same 576×576 pixel resolution. White pixels in the masks correspond to lesion localization. Several examples from the training set are given in Fig.1, where the first row corresponds to images without pathology, the second one to images with several AD lesions in every image, and the last row contains masks that correspond to the pathology images from the second row. In the dataset each image contains up to 6 lesions and their distribution is shown in Fig.3 (left). As shown, the most images contain only 1 lesion. In addition, Fig.3 (right) shows distribution of AD lesion areas that reach the maximum of approximately 12,000 pixels with the median value of 1,648 pixels.
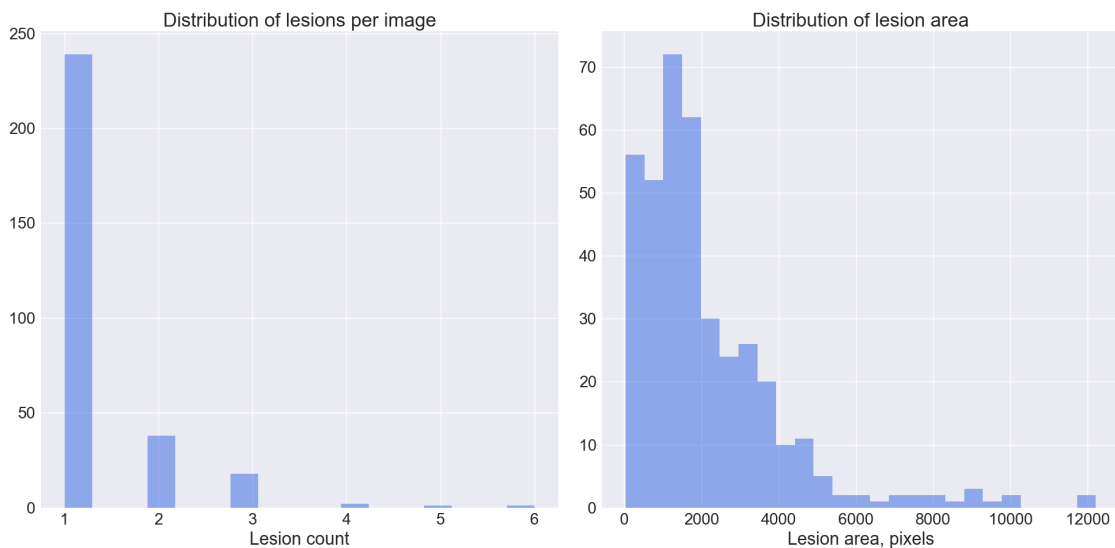

Figure 3: Distribution of angiodysplasia lesions per image (left figure) and distribution of lesions area (right figure) in the data set.

Images are cropped from 576×576 to 512×512 pixels to remove the canvas and text annotations. Then we rescale the data from $[0...255]$ to $[0...1]$ and standardize it following the ImageNet scheme (Iglovikov and Shvets, 2018). For training and cross-validation we only use 299 images annotated with binary masks that contain pathology. With those, we randomly split the dataset into five folds of 60, 60, 60, 60, and 59 images. In order to improve model generalization during training, random affine transformations and color augmentations in HSV space are applied.

2.2 Model architecture and training

In this work we evaluate 4 different deep architectures for segmentation: U-Net (Ronneberger et al., 2015; Iglovikov et al., 2017a), 2 modifications of TernausNet (Iglovikov and Shvets, 2018), and AlbuNet34, a modifications of LinkedNet (Chaurasia and Culurciello, 2017; Shvets et al., 2018).

In general, a U-Net-like architecture consists of a contracting path to capture context and of a symmetrically expanding path that enables precise localization (for example, see Fig.4). The contracting path follows the typical architecture of a convolutional network with alternating convolution and pooling operations and progressively downsamples feature maps, increasing the number of feature maps per layer at the same time. Every step in the expansive path consists of an upsampling of the feature map followed by a convolution. Hence, the expansive branch increases the resolution of the output. In order to localize, upsampled features, the expansive path combines them with high-resolution features from the contracting path via skip-connections (Ronneberger et al., 2015). The output of the model is a pixel-by-pixel mask that shows the class of each pixel. We use slightly modified version of the original U-Net model that previously proved itself very useful for segmentation problems with limited amounts of data, for example, see (Iglovikov et al., 2017a,b). Our winning submission to the MICCAI 2017 Endoscopic Vision SubChallenge: Angiodysplasia detection and localization (MICCAI 2017 Endoscopic Vision Challenge) was produced using this architecture.

As an improvement over the standard U-Net architecture, we use similar networks with pre-trained encoders. TernausNet (Iglovikov and Shvets, 2018) is a U-Net-like architecture that uses relatively simple pre-trained VGG11 or VGG16 (Simonyan and Zisserman, 2014) networks as an encoder (see Fig. 4). VGG11 consists of seven convolutional layers, each followed by a ReLU activation function, and five max polling operations, each reducing feature map by 2. All convolutional layers have 3×3 kernels. TernausNet16 has a similar structure and uses VGG16 network as an encoder (see Fig.4).

In contrast, AlbuNet uses an encoder based on a ResNet-type architecture (He et al., 2016). In this work, we use pre-trained ResNet34, see Fig.5. The encoder starts with the initial block that performs convolution with a kernel of size 7×7 and stride 2. This block is followed by max-pooling with stride 2. The later portion of the network consists of repetitive residual blocks. In every residual block, the first convolution operation is implemented with stride 2 to provide downsampling, while the rest convolution operations use stride 1. In addition, the decoder of the network consists of several decoder blocks that are connected with the corresponding encoder block. In this case, the transmitted block from the encoder is added to the corresponding decoder block. Each decoder block includes 1×1 convolution
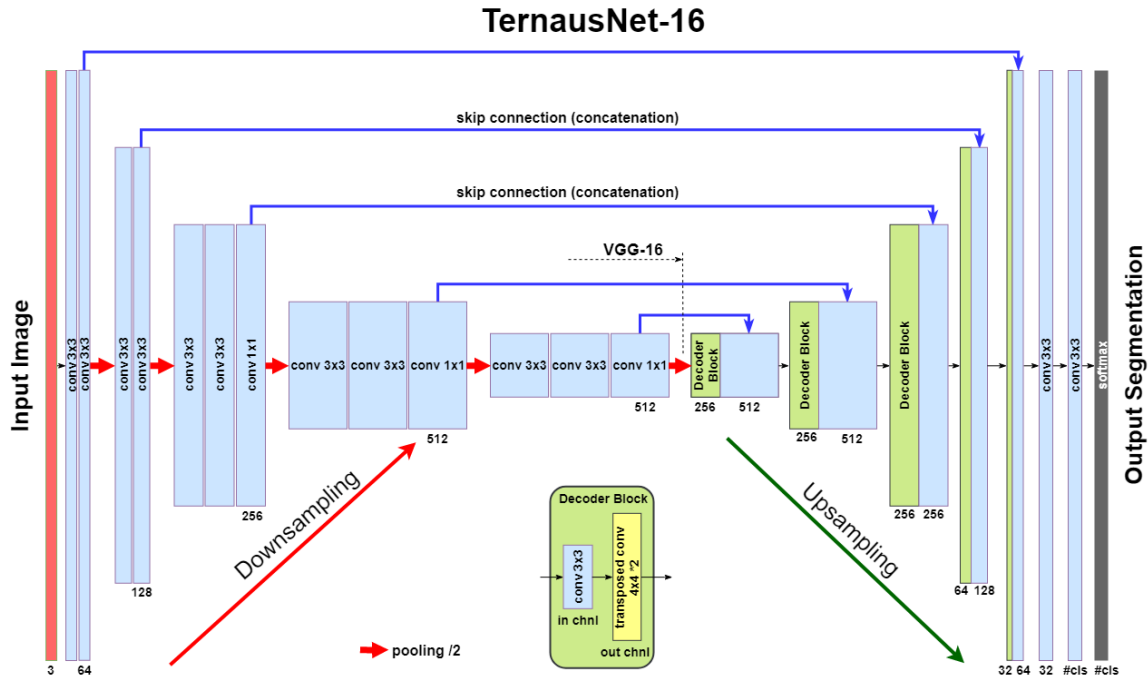
Figure 4: Segmentation networks based on encoder-decoder architecture of U-Net family. TernausNet uses pre-trained VGG16 network as an encoder Each box corresponds to a multi-channel feature map. The number of channels is pointed below the box. The height of the box represents a feature map resolution. The blue arrows denote skip-connections where information is transmitted from the encoder to the decoder.

operation that reduces the number of filters by 4, followed by batch normalization and transposed convolution to upsample the feature map.

We use Jaccard index (Intersection Over Union) as the evaluation metric. It can be interpreted as a similarity measure between a finite number of sets. For two sets A and B, it can be defined as following:

$$J(A, B) = \frac{|A \cap B|}{|A \cup B|} = \frac{|A \cap B|}{|A| + |B| - |A \cap B|} \tag{1}$$

Since an image consists of pixels, the last expression can be adapted for discrete objects in the following way:

$$J = \frac{1}{n} \sum_{i=1}^{n} \left(\frac{y_i \hat{y}_i}{y_i + \hat{y}_i - y_i \hat{y}_i} \right) \tag{2}$$

where y_i and $\hat{y}_i$ are a binary value (label) and a predicted probability for the pixel i, correspondingly.

Since image segmentation task can also be considered as a pixel classification problem, we additionally use common classification loss functions, denoted as H. For a binary seg-

mentation problem H is a binary cross entropy, while for a multi-class segmentation problem H is a categorical cross entropy.

The final expression for the generalized loss function is obtained by combining (2) and H as following:

$$L = H - \log J \tag{3}$$

By minimizing this loss function, we simultaneously maximize probabilities for right pixels to be predicted and maximize the intersection J between masks and corresponding predictions. We refer reader to Iglovikov et al. (2017a) for further details. Each model is trained with Adam optimizer (Kingma and Ba, 2014) for 10 epochs with learning rate 0.001, and then for another 5 epochs with the learning rate 0.0001.

As an output of a model, we obtain an image, in which each pixel value corresponds to a probability of belonging to the area of interest or a class. The size of the output image matches the input image size. For binary segmentation, we use 0.3 as a threshold value (chosen using validation dataset) to binarize pixel probabilities. All pixel values below the specified threshold are set to 0, while all values above the threshold are set to 255 to produce final prediction mask.

Following the segmentation step, we perform postprocessing in order to find the coordinates of angiodysplasia lesions in the image. In the postprocessing step we use OpenCV
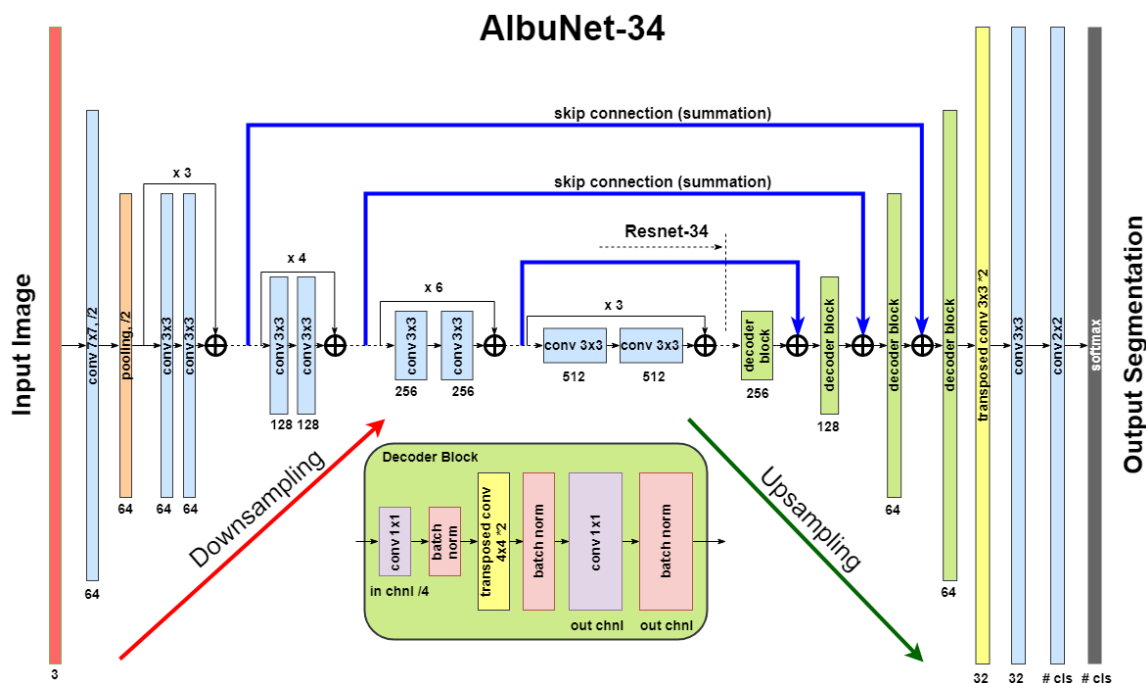

Figure 5: AlbuNet-34 uses pre-trained ResNet-34 as an encoder. It is different from TernausNet in that is adds skip-connections to the upsampling path, while TernausNet concatenates downsampled layers with the upsampling path (just like original U-Net does)

7

implementation of connected component labeling function: connectedComponentsWith-Stats (Bradski, 2000). This function returns the number of connected components, their sizes (areas), and centroid coordinates of the corresponding connected component. In our detector we use another threshold to neglect all clusters with the size smaller than 300 pixels. Therefore, in order to establish the presence of the lesions, the number of found components should be higher than 0, otherwise the image corresponds to a normal condition. Then, for localization of angiodysplasia lesions we return centroid coordinates of all connected components.

3. Results

To test our prediction and compare it with known mask we performed calculations on an image taken from the validation set. The exemplar result of the prediction is shown in Fig.6. For a visual comparison we also provide the original image and its corresponding mask. Given imperfect segmentation, this example does show that the algorithm sucessfully detects angiodysplasia lesions. When there are few lesions in an image and they are well separated in space, the detector performs almost very well. In case of many lesions that somehow overlap in space, further improvements are required, specifically in choosing model hyperparameters, to achieve better performance.

The quantitative comparison of our models' performance is presented in the Table1. For the segmentation task the best results is achieved by AlbuNet-34 providing $IoU = 0.754$ and $Dice = 0.831$. When compared by the inference time, AlbuNet is also the fastest model due to the light encoder. In the segmentation task this network takes around $20ms$
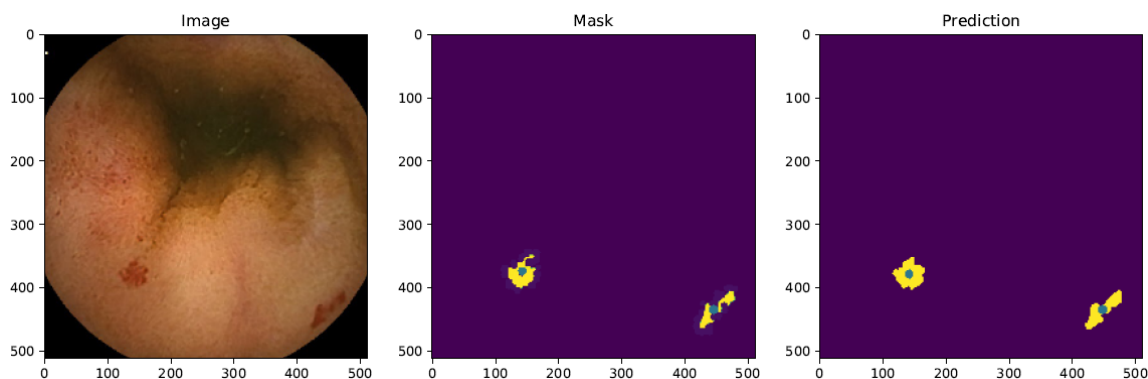

Figure 6: The prediction of our detector on the validation data image. Here, the first picture correspond to original image, the second one to the training mask, the last one to the predicted mask. Green dots inside the clusters corresponds to the centroid coordinates that define a localization of the appropriate angiodysplasia. For example, the real values for centroid coordinates are $p^1_{mask} = (376, 144)$, $p^1_{pred} = (380, 143)$ for the first cluster and $p^2_{mask} = (437, 445)$, $p^2_{pred} = (437, 447)$ for the second one.

Table 1: Segmentation results. Intersection over Union (IoU) and Dice coefficient (Dice) are in % and inference time (Time) is in ms.

Model	IOU	Dice	Time
U-Net	73.18	83.06	30
TernausNet-11	74.94	84.43	51
TernausNet-16	73.83	83.05	60
AlbuNet-34	**75.35**	**84.98**	**21**

for 512×512 pixel image and more than three times as fast as TernausNets. The inference time was measured using one NVIDIA GTX 1080Ti GPU.

4. Conclusions

We present deep learning-based segmentation and detection algorithms for angiodysplasia lesions localization in video capsule endoscopy. This study compares U-Net network architecture with its improved modifications that use custom pretrained encoders. Subsequent postprocessing based on the analysis of connected components is used to further refine predictions. Our approach shows quite good results on the validation. To the best of our knowledge, this study presents the first attempt in appllication of convolution neural networks for the problem of angiodysplasia lesion detection and classification. We demostrate state-of-the-art results in the MICCAI 2017 Endoscopic Vision SubChallenge: Angiodysplasia Detection and Localization. These results can be further improved by more accurate hyperparameter tuning as well as better postprocessing of connected components. Our code is available as an open source project under MIT licence at https://github.com/ternaus/angiodysplasia-segmentation.

References

Abbasi J Akhtar, Magda A Shaheen, and Junli Zha. Organic colonic lesions in patients with irritable bowel syndrome (ibs). *Medical science monitor*, 12(9):363–367, 2006.

G. Bradski. The OpenCV Library. *Dr. Dobb's Journal of Software Tools*, 2000.

Michael F Byrne, Nicolas Chapados, Florian Soudan, Clemens Oertel, Milagros Linares Pérez, Raymond Kelly, Nadeem Iqbal, Florent Chandelier, and Douglas K Rex. Real-time differentiation of adenomatous and hyperplastic diminutive colorectal polyps during analysis of unaltered videos of standard colonoscopy using a deep learning model. *Gut*, 2017. ISSN 0017-5749. doi: 10.1136/gutjnl-2017-314547. URL http://gut.bmj.com/content/early/2017/11/09/gutjnl-2017-314547.

Abhishek Chaurasia and Eugenio Culurciello. Linknet: Exploiting encoder representations for efficient semantic segmentation. *arXiv preprint arXiv:1707.03718*, 2017.

Travers Ching, Daniel S. Himmelstein, Brett K. Beaulieu-Jones, Alexandr A. Kalinin, Brian T. Do, Gregory P. Way, Enrico Ferrero, Paul-Michael Agapow, Michael Zietz,

Michael M. Hoffman, Wei Xie, Gail L. Rosen, Benjamin J. Lengerich, Johnny Israeli, Jack Lanchantin, Stephen Woloszynek, Anne E. Carpenter, Avanti Shrikumar, Jinbo Xu, Evan M. Cofer, Christopher A. Lavender, Srinivas C. Turaga, Amr M. Alexandari, Zhiyong Lu, David J. Harris, Dave DeCaprio, Yanjun Qi, Anshul Kundaje, Yifan Peng, Laura K. Wiley, Marwin H. S. Segler, Simina M. Boca, S. Joshua Swamidass, Austin Huang, Anthony Gitter, and Casey S. Greene. Opportunities and obstacles for deep learning in biology and medicine. *Journal of The Royal Society Interface*, 15(141), 2018. doi: 10.1098/rsif.2017.0387.

P Gregory Foutch, Douglas K Rex, and David A Lieberman. Prevalence and natural history of colonic angiodysplasia among healthy asymptomatic people. *American Journal of Gastroenterology*, 90(4), 1995.

Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 770–778, 2016.

Dimitris K Iakovidis and Anastasios Koulaouzidis. Software for enhanced video capsule endoscopy: challenges for essential progress. *Nature Reviews Gastroenterology and Hepatology*, 12(3):172, 2015.

Vladimir Iglovikov and Alexey Shvets. Ternausnet: U-net with vgg11 encoder pre-trained on imagenet for image segmentation. *arXiv preprint arXiv:1801.05746*, 2018.

Vladimir Iglovikov, Sergey Mushinskiy, and Vladimir Osin. Satellite imagery feature detection using deep convolutional neural network: A kaggle competition. *arXiv preprint arXiv:1706.06169*, 2017a.

Vladimir Iglovikov, Alexander Rakhlin, Alexandr Kalinin, and Alexey Shvets. Pediatric bone age assessment using deep convolutional neural networks. *arXiv preprint arXiv:1712.05053*, 2017b.

Alexandr A Kalinin, Gerald A Higgins, Narathip Reamaroon, SM Soroushmehr, Ari Allyn-Feuer, Ivo D Dinov, Kayvan Najarian, and Brian D Athey. Deep learning in pharmacogenomics: From gene regulation to patient stratification. *arXiv preprint arXiv:1801.08570*, 2018.

Alexandros Karargyris and Nikolaos Bourbakis. Wireless capsule endoscopy and endoscopic imaging: A survey on various methodologies presented. *IEEE Engineering in medicine and biology magazine*, 29(1):72–83, 2010.

Diederik P. Kingma and Jimmy Ba. Adam: A method for stochastic optimization. *CoRR*, abs/1412.6980, 2014. URL http://arxiv.org/abs/1412.6980.

S Lecleire, I Iwanicki-Caron, A Di-Fiore, C Elie, R Alhameedi, S Ramirez, S Hervé, E Ben-Soussan, P Ducrotté, and M Antonietti. Yield and impact of emergency capsule enteroscopy in severe obscure-overt gastrointestinal bleeding. *Endoscopy*, 44(4):337, 2012.

Zhuan Liao, Rui Gao, Can Xu, and Zhao-Shen Li. Indications and detection, completion, and retention rates of small-bowel capsule endoscopy: a systematic review. *Gastrointestinal endoscopy*, 71(2):280–286, 2010.

M. Mackiewicz, J. Berens, and M. Fisher. Wireless capsule endoscopy color video segmentation. *IEEE Transactions on Medical Imaging*, 27(12):1769–1781, Dec 2008. ISSN 0278-0062. doi: 10.1109/TMI.2008.926061.

R Marmo, G Rotondano, R Piscopo, MA Bianco, and L Cipolletta. Meta-analysis: capsule enteroscopy vs. conventional modalities in diagnosis of small bowel diseases. *Alimentary pharmacology & therapeutics*, 22(7):595–604, 2005.

MICCAI 2017 Endoscopic Vision Challenge. MICCAI 2017 Endoscopic Vision Challenge: Angiodysplasia Detection and Localization. https://endovissub2017-giana.grand-challenge.org/angiodysplasia-etisdb/.

Daniel S Mishkin, Ram Chuttani, Joseph Croffie, James DiSario, Julia Liu, Raj Shah, Lehel Somogyi, William Tierney, Louis M Wong Kee Song, and Bret T Petersen. Asge technology status evaluation report: wireless capsule endoscopy. *Gastrointestinal endoscopy*, 63(4):539–545, 2006.

Venkatesh N Murthy, Vivek Singh, Shanhui Sun, Subhabrata Bhattacharya, Terrence Chen, and Dorin Comaniciu. Cascaded deep decision networks for classification of endoscopic images. In *Medical Imaging 2017: Image Processing*, volume 10133, page 101332B. International Society for Optics and Photonics, 2017.

Alexander Rakhlin, Alexey Shvets, Vladimir Iglovikov, and Alexandr A Kalinin. Deep convolutional neural networks for breast cancer histology image analysis. *arXiv preprint arXiv:1802.00752*, 2018.

Jaroslaw Regula, Ewa Wronska, and Jacek Pachlewski. Vascular lesions of the gastrointestinal tract. *Best Practice & Research Clinical Gastroenterology*, 22(2):313–328, 2008.

Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 234–241. Springer, 2015.

S. S. Sami, S. A. AlAraji, and K. Ragunath. Review article: gastrointestinal angiodysplasia pathogenesis, diagnosis and management. *Alimentary Pharmacology & Therapeutics*, 39 (1):15–34, 2014. doi: 10.1111/apt.12527. URL https://onlinelibrary.wiley.com/doi/abs/10.1111/apt.12527.

Alexey Shvets, Alexander Rakhlin, Alexandr A Kalinin, and Vladimir Iglovikov. Automatic instrument segmentation in robot-assisted surgery using deep learning. *arXiv preprint arXiv:1803.01207*, 2018.

R Sidhu, DS Sanders, AJ Morris, and ME McAlindon. Guidelines on small bowel enteroscopy and capsule endoscopy in adults. *Gut*, 57(1):125–136, 2008.

Karen Simonyan and Andrew Zisserman. Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*, 2014.

Piotr Szczypiski, Artur Klepaczko, Marek Pazurek, and Piotr Daniel. Texture and color based image segmentation and pathology detection in capsule endoscopy videos. *Computer Methods and Programs in Biomedicine*, 113(1):396 – 411, 2014. ISSN 0169-2607. doi: https://doi.org/10.1016/j.cmpb.2012.09.004. URL http://www.sciencedirect.com/science/article/pii/S0169260712002192.

Nima Tajbakhsh, Suryakanth R Gurudu, and Jianming Liang. Automatic polyp detection in colonoscopy videos using an ensemble of convolutional neural networks. In *Biomedical Imaging (ISBI), 2015 IEEE 12th International Symposium on*, pages 79–83. IEEE, 2015.

Aleksei Tiulpin, Jérôme Thevenot, Esa Rahtu, Petri Lehenkari, and Simo Saarakkala. Automatic knee osteoarthritis diagnosis from plain radiographs: A deep learning-based approach. *Scientific Reports*, 8:1727, 2018. doi: 10.1038/s41598-018-20132-7.

Stuart L Triester, Jonathan A Leighton, Grigoris I Leontiadis, David E Fleischer, Amy K Hara, Russell I Heigh, Arthur D Shiff, and Virender K Sharma. A meta-analysis of the yield of capsule endoscopy compared to other diagnostic modalities in patients with obscure gastrointestinal bleeding. *The American journal of gastroenterology*, 100(11): 2407, 2005.

Yixuan Yuan and Max Q-H Meng. Deep learning for polyp recognition in wireless capsule endoscopy images. *Medical physics*, 44(4):1379–1389, 2017.