Small Bowel Motility Assessment based on Attentive Network

Xing Wu*
Shanghai University
xingwu@shu.edu.com

Mingyu Zhong
Shanghai University
ms_nymph@shu.edu.com

Yike Guo
Shanghai University
y.guo@imperial.ac.uk

ABSTRACT

The small bowel is the longest part of the gastrointestinal tract and quick assessment of its motility with Cine-MRI is of importance in the diagnosis of gastroenteric diseases. Because of the complex shape change of small bowel at any time, approaches with human designed features and simple CNN methods both fail to achieve satisfactory performance on massive datasets. To meet the challenge of assessing small bowel motility automatically, deformable convolutional networks are integrated into attentive network. With the help of deformable convolution, the tailored CNN learns to select the region of interest in a MR image. The proposed attentive encoder-decoder evaluates the current slice of data in view of the whole MRI sequence with a global attention context vector. After taking a guide line for a Cine-MRI sequence, the proposed method automatically marks the diameters on the original MR images to assess the temporal fluctuation of diameter lengths of the small bowel. Experimental results demonstrate that the proposed system outperform state-of-art methods, which fits to the shape change of small bowel and can be utilized to assistant diagnosis of gastroenteric diseases.

CCS CONCEPTS

• **Applied computing → Life and medical sciences**; **Imaging**; *Computational biology*;

KEYWORDS

cine MRI, small bowel motility, recursive neural networks, attention mechanism, deformable convolution

ACM Reference Format:
Xing Wu, Mingyu Zhong, and Yike Guo. 1997. Small Bowel Motility Assessment based on Attentive Network. In *Proceedings of ACM Woodstock conference (KDD'18)*, Jennifer B. Sartor, Theo D'Hondt, and Wolfgang De Meuter (Eds.). ACM, New York, NY, USA, Article 4, 6 pages. https://doi.org/10.475/123_4

1 INTRODUCTION

Small bowel plays an important role in digestion and desorption. The motility function of small bowel is essential to mix and grind the content in the bowel and transport. [19] To assess the motility of small bowel, several radiologic methods such as cross-sectional tomographic (CT), ultrasonography (US) and Magnetic Resonance Imaging (MRI) are applicable. Depending on the way of observation, certain form of data, generally images, with rich information of the internal state of the small bowel, will be generated. Radiologists then observe such data to determine whether a patient is in danger of suffering gastroenteric disorders like Crohn's disease [13]. Researchers have been exploring automatic systems to replicate the work of human experts. Recently, with the development of deep learning, certain neural network architectures did replicate the work or even outperform practicing radiologist.[6, 7, 11, 18]. We estimate the diameter of small bowel by evaluating images captured by cine-MRI which is a non-invasive modality [16, 19] that allows us observing both intra-luminal and extra-luminal abnormalities.
Convolutional neural networks (CNNs) and recursive neural networks (RNNs) are broadly used in the researches above, and they are proved to be extraordinarily effective(e.g. [6, 17, 18]). CNNs are considerably powerful in extracting nonlinear, hierarchical features [10]. As a refined version of CNN, deformable convolutional networks enhance CNN's transformation modeling capability by applying additional offsets to its spatial sampling locations[14].
Encoder-decoder is proposed first in the region of Neural Machine Translation (NMT)[3]. The system reads all words of a sentence written in original language, and emitting one translated word at a time to complete the translation process. Bahdanau et al. (2015) then applied attentional mechanism to jointly translate and align words[1].
We proposed an attentive network architecture that integrates deformable convolutional network and attentive encoder-decoder. In this approach, deformable convolutional network learns to select a region of interest of a MRI frame and extract high level features, while the attentive encoder-decoder takes those features, and learns to evaluate the current slice of data in consideration of the whole sequence with a global attention context vector. The criterion of prediction is the mean square error between predicted diameters and ground truth, therefore, Principle Components Analysis (PCA), a dimension reduction method, is applied on the output images sequence in order to calculate the the diameter length of the line marked on 2D image. PCA transforms the predicted 2D diameter into a 1D pattern, then the diameter length is obtained by measuring the length of this 1D representation.
Figure 1 illustrates an overview of the proposed the method.

2 RELATED WORK

In this section we report relevant background on some previous work on automatic and semi-automated small bowel assessment.
Earlier, some researchers[5, 16, 19] conducted manunal measurements on small bowel's diameters. Even with the help of guide-lines, such process is still both labor-intensive and time-consuming. In order to tackle with this problem, researches such as [13, 21] start to use quantitative techniques to complete the assessment. With

*Dr. Wu is the corresponding author of this article.

Permission to make digital or hard copies of part or all of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for third-party components of this work must be honored. For all other uses, contact the owner/author(s).
KDD'18, July 1997, El Paso, Texas USA
© 2016 Copyright held by the owner/author(s).
ACM ISBN 123-4567-24-567/08/06…$15.00
https://doi.org/10.475/123_4

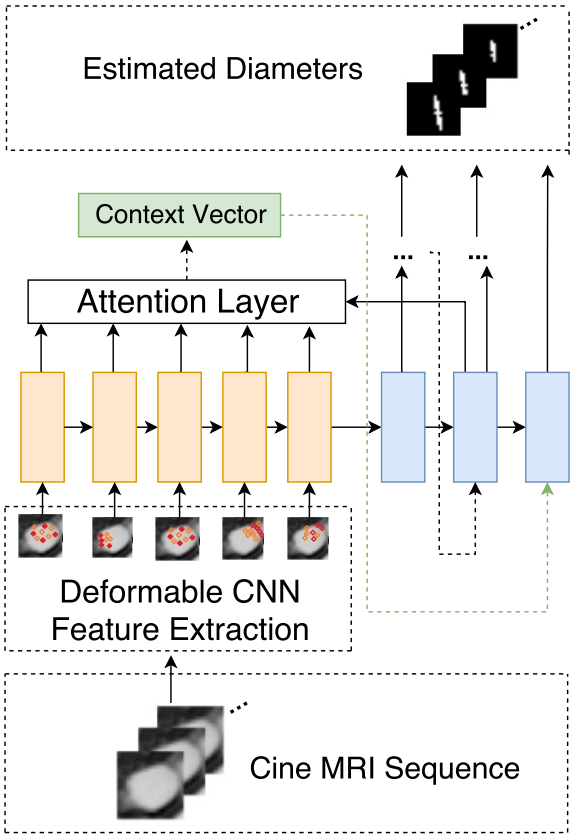

Figure 1: Overview of our system
The deformable convolutional network has attention on the image, while the encoder-decoder, with a attention layer, is attentive sequence-wise.

joint nonrigid registration and level set with automatic initialization introduced, researchers now regard the assessment of small bowel as a computer vision problem. Elaborate features are then specially designed to delineate luminal diameters automatically. However, these features are highly domain dependent, and can be difficult to capture.

With the ability to autonomously learn to extract features, deep neural networks have been introduced to medical image segmentation in a considerable number of studies[6, 7, 11, 17, 18]. Notablely, Rajpurkar et al. developed a 121-layer convolutional neural that can detect pneumonia from chest X-rays at a level exceeding practicing radiologists. Pei et al. proposed FCN-LSTM, an automated approach to mark the diameter of the bowel loop in every frame of the input sequence[17]. The work confirmed that CNNs and LSTMs, a variant of RNNs can be successfully applied to estimate the diameters of small bowel. However, it is also reported in the paper that the performance is still not satisfying.

Table 1 shows the comparison between these approaches briefly.

In the context of NMT, Cho et al.[3] proposed encoder-decoder, a novel system to learn sequence to sequence mapping, and [1, 12] introduced attention mechanism to it. Researches like [8, 14] proposed

Table 1: Results of previous approaches

Method	Authors	Result
Manual	Wakamiya et.al.[19]	*apm* : 10.4 ± 4.53mm
	Patak et.al.[16]	*apm* : 10 ± 4.7mm
Semi-automated	Froehlich et.al.[5]	*amp* : 6.55 ± 1.15
	Wu et.al.[21]	95% accuracy
Automated	Odille et.al.[13]	93%*accuracy*
	Pei et.al.[17]	MSE:4.472

quantitative adjustment on convolutional network to enhance its feature extracting ability. With such methods CNNs become *attentive* and compatible with dilatations and rotation.

The cine MRI images are exactly rich in local dilatation patterns, and the assessment can be view as a sequence to sequence mapping problem. However, there is no investigation of how these mechanism work on small bowel assessment. We integrated deformable convolutional networks into an encoder-decoder to put these mechanism into the small bowel assessment task, and verified that they outperform state-of-art methods.

3 THE ATTENTIVE NETWORK

In this section, we describe the attentive networks first by introducting the deformable CNN and attentive encoder-decoder. After that we illustrate how we combine these components together by giving a pseudocode snippet. Along with the equations that define the prediction and training process of our system, we will also give some more details on why our system is designed this particular way.

3.1 Deformable Convolutional Networks

Typically, CNNs are used to extract hierarchical features, the extracted vectorial representations can be directly used to conduct classification[10] or even caption generation[22]. However, the subsampling of CNN is achieved by applying rectangle convolutional kernels on the input and feature map, which prevent the network to learn complex unknown transformations. It is illustrated in the following figure that MRI frames are rich in local dilatations. We assume that it is proper to inffer that dynamic convolutional kernels help to extract features out of small bowel MRI.

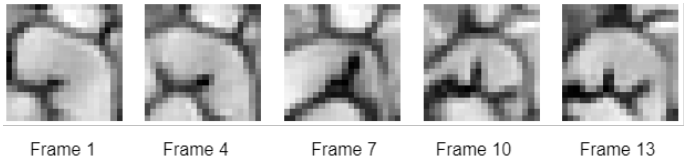

Figure 2: MRI images sequence

Concretely, given a sequence S_i, S_{ij} is one frame in S_i at time j, F_i denotes the extracted feature sequence, CNNs as feature extractor, learn the mapping

$$M_1 : S_i \rightarrow F_i \tag{1}$$

in the form of a series of convolutional layers.
In vanilla convolutional layers, convolution is applied

$$y(p_{i+1}) = \sum_{o} w(o) * x(p_i + o) \tag{2}$$

where $y(p_{i+1})$ is the value of point p_{i+1} on feature map, o(as offset) enumerate the deviation between each points in the convolution kernel and a point p_i in the input x, which is also in the kernel. Not quite like normal CNNs, the deformable convolutional networks train kernels with adaptive scales, by additional offset

$$y(p_{i+1}) = \sum_{o} w(o) * x(p_i + o + \Delta o) \tag{3}$$

As Δo may not be intergers, bilinear interpolation is introduced to implement (3). We combine such dilatation of offset in convolutional kernel into a full convolutional (FC) block in order to obtain better features, the network is illustrated in figure 3.

The first convolutional layer has 32 3x3 kernels and stride 1, zero-padding is performed to maintain the size of feature map. The second has the same configuration of kernels but has stride = 2. Both convolutional layers are deformable. We did not include any maxpooling layer.
Subsequently, two upsampling layers reform the feature map, and followed by a 1x1 normal convolutional layers. The last layer is meant to adjust the final output feature, hence there is no softmax layer at last.
Now with feature extracted, for a sequence S_i of length L, we have $F_i = \{F_{i1}, F_{i2}, ..., F_{iL}\}$. In the next subsection, we introduce the encoder-decoder that takes the feature sequence F_i, and generates diameters prediction Y_i.

3.2 Attentive Encoder Decoder

As illustrated in figure 1, our system has an encoder-decoder part that takes a sequence of extracted features and generates a sequences of predicted diameter endpoint images.

$$M_2 : F_i \rightarrow Y_i \tag{4}$$

We closely follows the works of Cho et al.[2], uses GRUs, a variant of RNNs, to make each recurrent unit to adaptively capture dependencies of different time scales[4]. Both encoder and decoder is one layer of GRU units.
At time step t corresponding to frame j, the encoder takes a feature input F_{ij}, and save a hidden state eh_{ij}, for a sequence of feature inputs F_i we have corresponding encoder hidden state sequence eh_i.

$$eh_{i,j+1} = (1 - z_j)eh_{i,j} + z_j\tilde{h}_j \tag{5}$$

where an update gate z_j controls how we generate $eh_{i,j+1}$ by applying linear interpolation between previous hidden state $eh_{i,j}$ and candidate state $\tilde{h}_j$.
By introducing a reset gate r_j, z_j, $\tilde{h}_j$ and r_j can be computed by

$$z_j = \sigma(W_z F_{ij} + U_z h_{j-1}) \tag{6}$$

$$\tilde{h}_j = tanh(W F_{ij} + U(r_t \odot h_{t-1})) \tag{7}$$

$$r_j = \sigma(W_r F_{ij} + U_r h_{j-1}) \tag{8}$$

The decoder produces a diameter prediction conditioned on the previous hidden state, the last generated result, and in addition, a context vector, where *attention* is achieved.
Originally in an NMT problem, the attention mechanism is used to align words of the source sentence and the target sentence. In addition to the normal version of RNN unit input, a context vector c is introduced, in order to get a representation of relationships between words. Xu et al.[22] successfully apply this technique to detect *where* an object is in a image. We are inspired by the works above. For each frame position j in the input feature sequence F_i, a positive weight a is generated, it could be interpreted as the significance the frame has, given the task to predicted a diameter at time t. In another word, the attention mechanism helps the decoder to decide how relative the current frame is to other frames in the sequence. More over, the properties of our data also suggest that a sequence level of attention might help. As illustrated in figure 4, a typical sequence of MRI contains multiple loops of constriction and eclasis, and the decoder will be able to take other loops as references with the help of context vector.

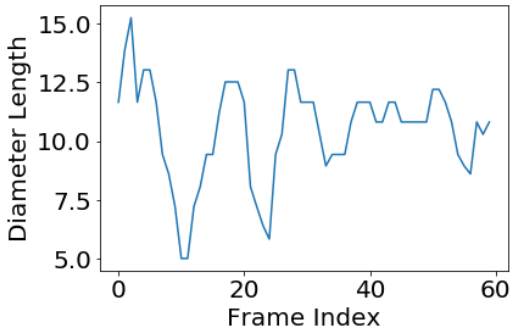

Figure 4: Diameter line chart of a typical MRI sequence

With the previous hidden state h_{t-1} at time t, that is, the decoder is prepared to predict the t th frame of the result sequence, an *energy* score is first calculated, with h_{t-1} and eh_i:

$$energy_t = eh_i^\top W h_{t-1} \tag{9}$$

where W is a parameter matrix. Note that $energy_t$ is a L dimensional vector, that is, it has exactly the same length L of eh_i. We then apply softmax to obtain the attention weight a_t

$$a_{tj} = softmax(energy_t) \tag{10}$$

$$= \frac{energy_{tj}}{\sum_{j=1}^{L} energy_{tj}} \tag{11}$$

with the weight we compute the context vector c_t

$$c_t = \sum_{j=1}^{L} a_{tj} eh_{ij} \tag{12}$$

and finally the output Y_i

$$Y_{i,j+1} = GRU(h_{t-1}, Y_{ij}, c_t) \tag{13}$$

3.3 Training and other details

With all essential components discribe above, in this subsection we will restate the prediction process, the training procedure, and

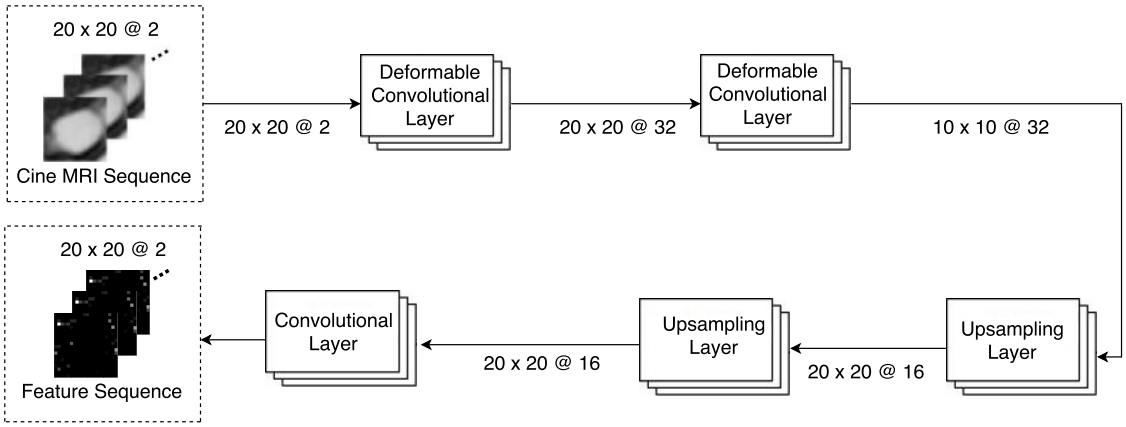

Figure 3: Demonstration of Deformable Feature Extraction

provide some detailed data flow in our system. A pseudocode is given at the end the subsection.

To recap, we have two mapping in the system

$$M_1 : S_i \rightarrow F_i$$
$$M_2 : F_i \rightarrow Y_i$$

Here, S_i is a sequence 2 channels images, obtained by stack an original MRI frame and a guide-line image together, we denote this relationship as $S_i = \{X_i, G_i\}$, where X_i is the original image sequence, and G_i is the guide-line sequence.

Apart from it, though we display predicted diameters in 2D form, we have flatten the image, as a result, Y_{ij} is a wh dimensional vector, where w and h denotes the width and height of the output image. Similarly, F_{ij} will be flatten to whc dimension before M_2 is conducted, where c is the number of channels of the feature map. At the last stage we will transform Y_{ij} back into 2D for human inspection and other quantitative operation like PCA.

Moreover, all the convolutional layers, deconvolutional layers and RNNs use ReLu function as activation function. It is verified that network with ReLu function works particularly well [10].

The system as a whole is end-to-end, with the MRI sequence and guide-lines, the result of prediction is directly obtained. Also, the network connection is smooth, differentiable at every state. As a result, using standard back-propagation to learn the parameters is trivial.

We define $S = \{S_1, S_2, ..., S_n\}$, $Y = \{Y_1, Y_2, ..., Y_n\}$, the ground truth set $\hat{Y} = \{\hat{Y}_1, \hat{Y}_2, ..., \hat{Y}_n\}$, using k to iterate wh and Y_{ijk} to denotes the k th pixel in frame j of sequence i, Θ as the parameters of whole system, and loss function as a pixel level mean square error(MSE)

$$loss(S, \hat{Y}, \Theta) = MSE(Y, \hat{Y})$$
$$= \frac{1}{whnL} \sum_{i=1}^{n} \sum_{j=1}^{L} \sum_{k=1}^{wh} (Y_{ijk} - \hat{Y}_{ijk})^2$$

During the training procedure, we calculate the partial derivatives with back-propagation and the learning algorithm optimizes the empirical loss.

However, no previous approach use pixel level loss, and makes the comparison between our results and them difficult, as a result, instead of updating the system with minimal empirical loss value, we select another criterion, using the MSE of predicted diameter length and real diameter length.In another word, when to evaluate the system output Y_i, it is first reshaped into a 20 x 20 2D images, then PCA transforms it back to 1D again, where diameter length D_i can be easily calculated. PCA is also conducted on ground truth $\hat{Y}_i$ to get $\hat{D}_i$. The back-propagation algorithm reduce our loss function in training loop, at the end of every epoch, we exam the MSE of D_i and $\hat{D}_i$. Like the definition above, we have $D = \{D_1, D_2, ..., D_n\}$, $\hat{D} = \{\hat{D_1}, \hat{D_2}, ..., \hat{D_n}\}$,

$$criterion = MSE(D, \hat{D})$$
$$= \frac{1}{whnL} \sum_{i=1}^{n} \sum_{j=1}^{L} \sum_{k=1}^{wh} (D_{ijk} - \hat{D}_{ijk})^2$$

If the criterion decreases, we update the system parameter Θ. In 4.2, we will show that although this choice of system updating ignores the position information of predicted diameters, it didn't bring about undesired side-affects on our results.

To sum up, we formally presented the training procedure of our system in Algorithm 1.

Algorithm 1 Training Attentive Network

initialize system Θ_b; $\Theta \leftarrow \Theta_b$
for each epoch **do**
  $F \leftarrow M_1(S)$; $F \leftarrow flatten(F)$; $Y \leftarrow M_2(F)$
  optimize loss function:
  $\Theta \leftarrow \Theta - lr * \nabla_\Theta \frac{1}{whnL} \sum_{i=1}^{n} \sum_{j=1}^{L} \sum_{k=1}^{wh} (Y_{ijk} - \hat{Y}_{ijk})^2$
  $D \leftarrow PCA(reshape(Y))$; $\hat{D} \leftarrow PCA(reshape(\hat{Y}))$
  if $criterion = MSE(D, \hat{D})$ decreases, **then**
    adapt new best system:
    $\Theta_b \leftarrow \Theta$
  end if
end for

Table 2: Comparison between previous approaches and our systems

Method	mean ± std	MSE
Manual[19]	10.4 ± 4.53	-
FCN-LSTM[17]	11.235 ± 2.032	4.472
CNN + Encoder-Decoder	8.87 ± 2.63	4.18
CNN + Attentive	8.86 ± 2.25	4.20
Deformable + Encoder-Decoder	8.86 ± 2.84	**3.34**
Deformable + Attentive	9.13 ± 2.62	**3.27**

4 EXPERIMENTS AND RESULT

4.1 Experiments

We conducted the experiments on a cine MRI database with 300 sequences, each consists of 60 images. 9 healthy volunteers without gastroenteric diseases participated the experiment. After 8 hours of fasting with transoral administration of 1500mL of nonabsorbable liquid prior to scanning, cine MRI is performed. Detailed MR machine parameters are listed in appendix A. ROIs and diameters are manunally selected and labeled by a radiologist. Each image sequence shares a guide-line image. The database are split randomly to a training set and a testing set with splitting ratio 5:1.

We conducted experiments on a deep learning server with 18x Intel(R) Xeon(R) CPU E5-2640 v4 @ 2.40GHz, 128G RAM, NVIDIA Tesla M40 24GB GPU enabled.

The attentive network is implemented based on the deep learning library PyTorch [15].

During the training procedure, a Adam optimizer[9] with betas (0.9, 0.999) recommended by the original authors, learning rate of 0.00001 is applied to optimize the loss function. Mini-batch size is set to 1, that is, in every mini-batch loop we use one sequence only and apply stochastic gradient descent. As discussed in 3.3, we terminate the training process after the MSE criterion converged.

To support that the attention opponents, namely deformable convolution and context vector help to accomplish the assessment task, we conduct our experiments on 4 systems

- Normal CNN + Normal Encoder-Decoder
- Normal CNN + Attentive Encoder-Decoder
- Deformable CNN + Normal Encoder-Decoder
- Deformable CNN + Attentive Encoder-Decoder

Results are discussed in the next subsection.

4.2 Result and Discussion

In Figure 5, the estimated diameters(masked on original images) and corresponding ground truth are displayed. It can be observed through these examples that the system can give well fitted patterns both for orientation and the length, at least at a visual level. Radiologists can readily exploit the results to conduct pathologic diagnosis and discard the arduous frame by frame diameter marking process. Moreover, quantitative analysis also suggests that our system well captured the segmental motion patterns. As stated in section 1 and subsection 3.3, we apply PCA on the 2D images to measure the euclidean distance between the endpoints of the diameter, and use the MSE as a criterion.

Table 2 displays our results and two of other approaches using similar criteria, including diameter prediction in a *mean* ± *std* fashion and the mean square error value. It is clear that the results of our system are within reasonable range while smaller error is achieved. In figure 6, we plot some cross-sectional attention weights within the prediction process.

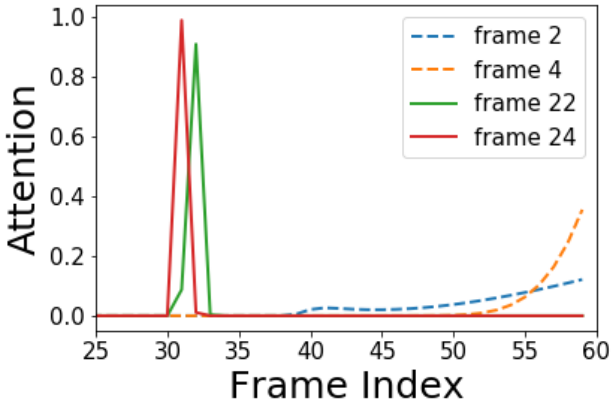

Figure 6: Attention distribution learned by the system
The x axis indicates index of frames 1-60, and the y axis indicates the relative proportion, weighting from 0 to 1. The omitted frames 0-25 have almost zero attention.

We found that at the begining of the prediction, typically between frame 1 to 10, the a value of the tail part of the sequence is higher, producing curves similar to the dotted lines in figure 6, later on the peak appears mainly on the middle part of the sequence, resulting curves similar to the full lines. As explained in section 3.2, we can interprete a as the a significance value, so it is safe to say, the network first gives more *attention* to the end of the MRI sequence, and then focuses on some frame in the middle. For instance, in the sequence we displays in figure 6, 41 out of 60 frames give the highest attention to frame 31.

We further presents frames near the 31th frame of the above sequence for result illustration in figure 7.

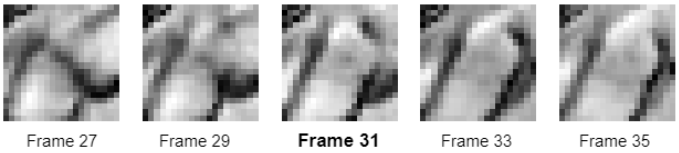

Figure 7: Frames near the key frame selected by the system

It can be seen from figure 7 that the frame selected is precisely the *average* one, right in the middle of the eclasis process. We can draw a preliminary conclusion that the attentive network is able to capture the crucial and representative portion of a small bowel MRI sequence.

Unfortunately, apart from preliminary result, the attention distribution learned by our system is not completely intuitively comprehensible. The reason why the system focuses on the tail part of

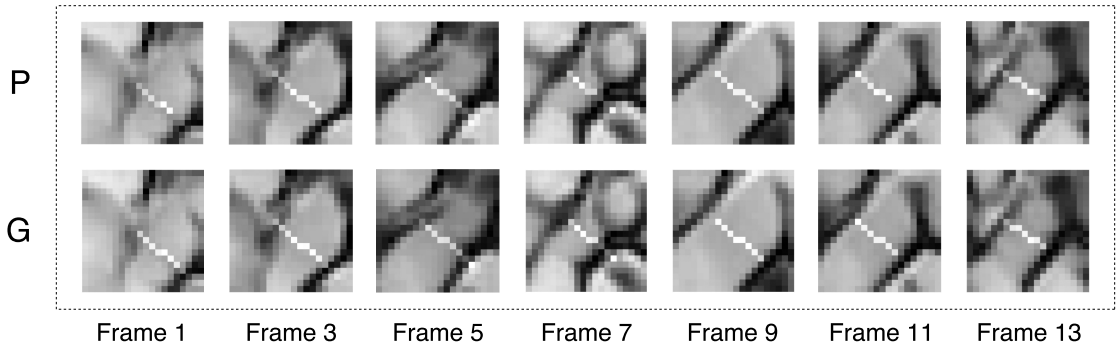

Figure 5: Examples of predicted diameters
P stands for predicted, G stands for ground truth.

the sequence at the begining phase remains unclear. More discussion with specialists in this field is needed if we are to determine whether the network treats representative parts of the sequence in the same way a human radiologist does.

A MRI MACHINE SETTINGS

We used a 1.5-T MR machine in this study. The parameter settings are identical with [20], we will quote them here:
Cine-MR imaging was performed with 1.5-T MR machine using 12-channel body array coil.
The balanced steady-state free precession imaging, FIESTA sequence (TR/TE=3.4/1.2msec, Flip angle=75 degree, acquisition time per image=0.5 sec) was utilized and the area of 45cm x 45cm was imaged to cover the entire loops of the small bowel.
Ten mm-thick coronal images were obtained at every 0.5 seconds for 30 seconds during breath hold at 0, 15, 30 ,45, and 60 minutes after oral intake of contrast.

ACKNOWLEDGMENTS

The authors would like to thank Mr. Pei for providing his code and detailed explanations of his work.

REFERENCES

[1] Dzmitry Bahdanau, Kyunghyun Cho, and Yoshua Bengio. 2014. Neural machine translation by jointly learning to align and translate. *arXiv preprint arXiv:1409.0473* (2014).
[2] Kyunghyun Cho, Bart Van Merriënboer, Dzmitry Bahdanau, and Yoshua Bengio. 2014. On the properties of neural machine translation: Encoder-decoder approaches. *arXiv preprint arXiv:1409.1259* (2014).
[3] Kyunghyun Cho, Bart Van Merriënboer, Caglar Gulcehre, Dzmitry Bahdanau, Fethi Bougares, Holger Schwenk, and Yoshua Bengio. 2014. Learning phrase representations using RNN encoder-decoder for statistical machine translation. *arXiv preprint arXiv:1406.1078* (2014).
[4] Junyoung Chung, Caglar Gulcehre, KyungHyun Cho, and Yoshua Bengio. 2014. Empirical evaluation of gated recurrent neural networks on sequence modeling. *arXiv preprint arXiv:1412.3555* (2014).
[5] Johannes M Froehlich, Michael A Patak, Constantin von Weymarn, Christoph F Juli, Christoph L Zollikofer, and Klaus-Ulrich Wentz. 2005. Small bowel motility assessment with magnetic resonance imaging. *Journal of Magnetic Resonance Imaging* 21, 4 (2005), 370–375.
[6] Monika Grewal, Muktabh Mayank Srivastava, Pulkit Kumar, and Srikrishna Varadarajan. 2017. RADNET: Radiologist Level Accuracy using Deep Learning for HEMORRHAGE detection in CT Scans. *arXiv preprint arXiv:1710.04934* (2017).
[7] Varun Gulshan, Lily Peng, Marc Coram, Martin C Stumpe, Derek Wu, Arunachalam Narayanaswamy, Subhashini Venugopalan, Kasumi Widner, Tom Madams, Jorge Cuadros, et al. 2016. Development and validation of a deep learning algorithm for detection of diabetic retinopathy in retinal fundus photographs. *Jama* 316, 22 (2016), 2402–2410.
[8] Max Jaderberg, Karen Simonyan, Andrew Zisserman, et al. 2015. Spatial transformer networks. In *Advances in Neural Information Processing Systems*. 2017–2025.
[9] Diederik P Kingma and Jimmy Ba. 2014. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980* (2014).
[10] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton. 2012. Imagenet classification with deep convolutional neural networks. In *Advances in neural information processing systems*. 1097–1105.
[11] Sancy A Leachman and Glenn Merlino. 2017. Medicine: The final frontier in cancer diagnosis. *Nature* 542, 7639 (2017), 36–38.
[12] Minh-Thang Luong, Hieu Pham, and Christopher D Manning. 2015. Effective approaches to attention-based neural machine translation. *arXiv preprint arXiv:1508.04025* (2015).
[13] Freddy Odille, Alex Menys, Asia Ahmed, Shonit Punwani, Stuart A Taylor, and David Atkinson. 2012. Quantitative assessment of small bowel motility by non-rigid registration of dynamic MR images. *Magnetic resonance in medicine* 68, 3 (2012), 783–793.
[14] Wanli Ouyang, Xiaogang Wang, Xingyu Zeng, Shi Qiu, Ping Luo, Yonglong Tian, Hongsheng Li, Shuo Yang, Zhe Wang, Chen-Change Loy, et al. 2015. Deepid-net: Deformable deep convolutional neural networks for object detection. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 2403–2412.
[15] Adam Paszke, Sam Gross, Soumith Chintala, Gregory Chanan, Edward Yang, Zachary DeVito, Zeming Lin, Alban Desmaison, Luca Antiga, and Adam Lerer. 2017. Automatic differentiation in PyTorch. (2017).
[16] Michael A Patak, Johannes M Froehlich, Constantin von Weymarn, Stefan Breitenstein, Christoph L Zollikofer, and Klaus-Ulrich Wentz. 2007. Non-invasive measurement of small-bowel motility by MRI after abdominal surgery. *Gut* 56, 7 (2007), 1023–1025.
[17] Mengqi Pei, Xing Wu, Yike Guo, and Hamido Fujita. 2017. Small bowel motility assessment based on fully convolutional networks and long short-term memory. *Knowledge-Based Systems* 121 (2017), 163–172.
[18] Pranav Rajpurkar, Jeremy Irvin, Kaylie Zhu, Brandon Yang, Hershel Mehta, Tony Duan, Daisy Ding, Aarti Bagul, Curtis Langlotz, Katie Shpanskaya, et al. 2017. CheXNet: Radiologist-Level Pneumonia Detection on Chest X-Rays with Deep Learning. *arXiv preprint arXiv:1711.05225* (2017).
[19] Makoto Wakamiya, Akira Furukawa, Shuzo Kanasaki, and Kiyoshi Murata. 2011. Assessment of small bowel motility function With cine-MRI using balanced steady-state free precession sequence. *Journal of Magnetic Resonance Imaging* 33, 5 (2011), 1235–1240.
[20] Xing Wu, Wu Zhang, and Qing Li. 2011. Quantitative Assessment of the Small Bowel Motility with Negative Mutual Information in Cine-MRI Sequences. *Procedia Environmental Sciences* 8 (2011), 328–336.
[21] Xing Wu, Shaojian Zhuo, and Wu Zhang. 2013. Automatic quantitative assessment of the small bowel motility with cine-MRI sequence analysis. In *International Symposium on Visual Computing*. Springer, 11–19.
[22] Kelvin Xu, Jimmy Ba, Ryan Kiros, Kyunghyun Cho, Aaron Courville, Ruslan Salakhudinov, Rich Zemel, and Yoshua Bengio. 2015. Show, attend and tell: Neural image caption generation with visual attention. In *International Conference on Machine Learning*. 2048–2057.