Attention U-Net: Looking Where to Look for the Pancreas

Raaghav Radhakrishnan[246097]

Information Systems and Machine Learning Lab,
Marienbürgerplatz 22, Universität Hildesheim, 31141 Hildesheim, Germany
radhakri@uni-hildesheim.de

Abstract. This paper reviews one of the interesting research directions that has emerged in image segmentation, medical image segmentation. Image segmentation as related to medical image processing is broadly used in the domain of Artificial Intelligence for cardiology, orthopedics and ophthalmology to name but a few. These applications help physicians diagnose major diseases in its earlier stages of development. As one of the research in medical applications, the paper proposes a new medical imaging attention gate (AG) model that automatically learns to concentrate on target organs of variable sizes and shapes. The AGs are integrated with minimal computational overhead into the standard U-Net model, while improving model sensitivity and predictive precision. For multi-class segmentation, the Attention U-Net architecture experimented on two CT abdominal datasets shows that attention gates improve the prediction performance to a smaller extent compared to the standard U-Net architecture. With the proposed approach, an external object localisation model is eliminated and also, the attention network can be initialized using the pre-trained U-Net weights.

Keywords: Medical image segmentation · U-Net · Attention gates · etc.

1 Introduction

This paper reviews the work “Attention U-Net: Learning Where to Look for the Pancreas” by O Oktay et al. It was published at the Conference on Medical Imaging with Deep Learning (MIDL 2018), Amsterdam, The Netherlands. The MIDL conference seeks to be a forum for deep learning researchers, clinicians and healthcare firms to take a step forward in applying deep learning-based automatic image analysis in disease screening, diagnosis, prognosis, treatment selection and treatment monitoring [16]. This research work is cited by top researchers like YoungJu Jo and Fabian Isensee and has a total of 45 citations [14].

The work on attention gated model is proposed by the authors - Ozan Oktay, Jo Schlemper, Loic Le Folgoc, Matthew Lee, Mattias Heinrich, Kazunari Misawa, Kensaku Mori, Steven McDonagh, Nils Y Hammerla, Bernhard Kainz, Ben Glocker and Daniel Rueckert. The information on some of the authors is as follows:

1. Ozan Oktay, PhD in Biomedical Image Analysis and Processing, is a Research Associate in the Department of Computing at Imperial College, London (ICL). His research focuses on medical image analysis for image superresolution, semantic image segmentation and object localisation. His works have a total of 794 citations [14].
2. Jo Schlemper is a PhD Student at BioMedIA, ICL. His interests include image reconstruction, compressed sensing and MRI. His research works have a total of 318 citations [14].
3. Loic Le Folgoc, PhD in Medical Imaging, is a Research Associate at BioMedIA, ICL. He is interested in signal processing and information geometry, computational biology and physics and agent-based models. His research works have a total of 141 citations [14].
4. Mattias Heinrich, PhD, is an Assistant Professor at Institute of Medical Informatics at University of Luebeck. His research focus lies in the development of deformable image registration tools. He has 1993 citation [14] in total for his research works.
5. Bernhard Kainz is an Assistant Professor in the Department of Computing at ICL. His research is related to interactive algorithms in healthcare, especially medical imaging. His publications have 1525 citations [14] in total.
6. Daniel Rueckert is a Professor and Head of the Department of Computing at ICl. He holds valuable positions in IEEE, MICCA/Elsevier Book Series and a referee for a number of international medical imaging journal and conferences. His works have a whopping citation of 37999 [14].

Due to the fact that manual annotation and labeling of huge amounts of medical pictures is a difficult and error-prone job, automated medical picture segmentation has been widely studied in the picture analysis community. Precise and reliable solutions are required to improve the effectiveness of clinical workflow and assist decision-making by quickly and automatically extracting quantitative measurements. In automated medical image analysis, with the emergence of convolutional neural networks (CNNs), it has been shown that a good performance can be achieved in the segmentation of cardiac MR [1] and detection of cancerous lung nodule [3].

In medical image segmentation, CNNs play a vital role because of its quick inference and high performing characteristics. The commonly used architectures, Fully Convolutional Networks (FCNs) and U-Net, use multi-stage cascaded CNNs to localize a region of interest and make predictions on the same. The problem with this approach is that the features of similar shape and size are repeatedly extracted by all the models within the cascade which makes the computation expensive. As proposed by the authors, without extra oversight, this problem can be solved by introducing AGs to the existing architecture that learns to focus on the target structures on its own. By suppressing the feature activation in irrelevant areas, AGs eliminate the need of large number of parameter in the multi-model frameworks. Hence, by introducing AGs, the authors claim that the prediction accuracy can be consistently enhanced without various CNN models being required.

The remainder of this review paper is organized as follows. Section 2 presents the related work and state-of-the-art. Section 3 summarizes the research work. Section 4 discusses the evaluation and review on the work. Section 5 concludes the discussion and its result.

2 Related Work

Although image segmentation in medical image analysis seems to be a common problem, lots of researches and experiments are being carried out to improve the performance of automated segmentation. The following are some of the researches and related works carried out for the problem.

In [12] and [6], the authors Wolz et al. and Oda et al. discuss the multi-atlas techniques for pancreas segmentation. Atlas approaches in specific benefit from implicit shape limitations that are implemented by manual annotation propagation.

A classification-based framework suggested by Zografos et al. in [13] removed the atlas dependence on image registration. Also, the problem is addressed using multi-stage CNN models in [8] and [9] by authors Roth et al. The proposed methods provide hands for automated pancreas localization and segmentation.

With an FCN [4] like U-Net [7], the authors Ronneberger et al. have showed that better and accurate performance can be achieved in various medical image segmentation tasks.

AGs are frequently used for image captioning, machine translation, and classification tasks [2, 11] in natural image analysis, knowledge graphs, and language processing. Of the two attentions, hard and soft, in [2] and [11], the authors Jetley et al. and Wang et al. have used soft attention for image classification. Also, the usage of these attentions resulted in improved performance on image classification.

3 Summary

A novel self-attention gating module proposed in this work can be used for dense label predictions in conventional image analysis models based on CNN. In addition, in the context of image segmentation in specific, the authors explore the benefits of AGs for medical image analysis. The work is summarized as follows:

1. The authors used the the attention mechanism proposed in [2] to improve the dense predictions of pancreas by giving importance to the relevant regions and suppressing the irrelevant ones.

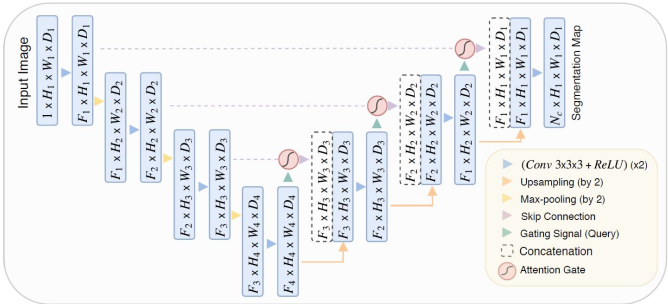

Input Image
$1 \times H_1 \times W_1 \times D_1$
$F_1 \times H_1 \times W_1 \times D_1$
$F_1 \times H_2 \times W_2 \times D_2$
$F_2 \times H_2 \times W_2 \times D_2$
$F_2 \times H_3 \times W_3 \times D_3$
$F_3 \times H_3 \times W_3 \times D_3$
$F_3 \times H_4 \times W_4 \times D_4$
$F_4 \times H_4 \times W_4 \times D_4$
$F_3 \times H_3 \times W_3 \times D_3$
$F_3 \times H_3 \times W_3 \times D_3$
$F_2 \times H_3 \times W_3 \times D_3$
$F_2 \times H_2 \times W_2 \times D_2$
$F_2 \times H_2 \times W_2 \times D_2$
$F_1 \times H_2 \times W_2 \times D_2$
$F_1 \times H_1 \times W_1 \times D_1$
$F_1 \times H_1 \times W_1 \times D_1$
$N_c \times H_1 \times W_1 \times D_1$
Segmentation Map
▶ $(Conv\ 3x3x3 + ReLU)$ (x2)
▶ Upsampling (by 2)
▶ Max-pooling (by 2)
▶ Skip Connection
▶ Gating Signal (Query)
⬚ Concatenation
Attention Gate

Fig. 1. Block diagram of Attention U-Net segmentation model

2. For the organ localization, the proposed soft-attention module not only replaced the object localization models but also replaced the existing hard-attention techniques.

3. As shown in Fig. 1), the standard U-Net model is extended to a step further by adding an attention to the skip connection layers and the gating layer in the bottle-neck.

3.1 Attention Gates

In standard CNN architectures, the feature maps are down scaled to capture the information at a coarser scale. This information can be useful when the maps are upscaled at a global scale. This objective can be successfully achieved by adding attention gates on top of a CNN model. Doing so, the training of different localisation models and other parameters can be eliminated. As shown in Fig. 2), the input features are scaled with the calculated attention coefficients. For each pixel a gating vector is used to determine areas of focus. The gating vector includes contextual information to prune feature responses at the lower level [11]. The attention mechanism is formulated as follows.

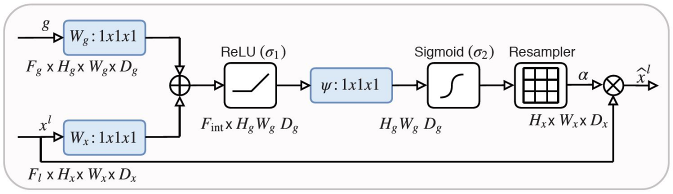


Fig. 2. Schematic of attention gate (AG)

$$q^l_{att} = \psi^T(\sigma_1(W^T_x x^l_i + W^T_g g_i + b_g)) + b_\psi \tag{1}$$

$$\alpha^l_i = \sigma_2((q^l_{att}((x^l_i, g_i; \theta_{att})), \tag{2}$$

where,
$\sigma_2(x_{i,c}) = \frac{1}{1+exp(-x_{i,c})}$ is the sigmoid activation function.
θ_{att} is a set of parameters containing weights of input (W_x) and gating (W_g) maps and bias terms ($b_\psi b_g$).
$\alpha_i \in [0, 1]$ is the attention coefficient for activating relevant regions.

Here, the attention coefficients extract the salient image regions to activate only the feature responses in the relevant regions. The Fig. 3), shows that the model gradually learns to focus on the pancreas, kidney and spleen. The output of AGs is formulated as follows:

$$\hat{x}^l_{i,c} = x^l_{i,c} * \alpha^l_i \tag{3}$$

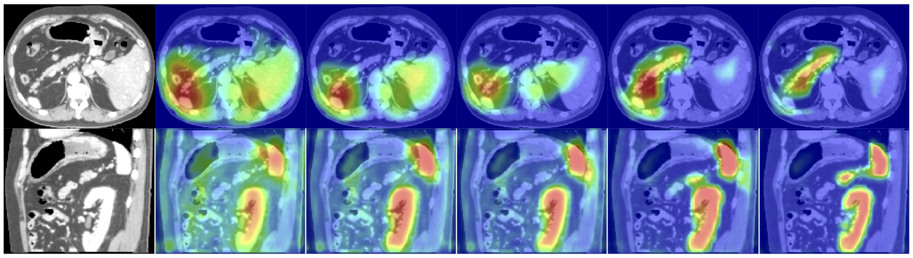


Fig. 3. Attention coefficients across different training epochs

3.2 Attention U-Net

The Attention U-Net architecture is an extended version of U-Net architecture with attention gates. The U-Net architecture is built on the Fully Convolutional Network and modified to enhance medical imaging segmentation. The difference between U-Net and Attention U-net is just the introduction of attention gates in the skip connections. The schematic representation of the Attention U-Net architecture is as follows. As shown in Fig. 4), the Attention U-Net architecture has 4 main phases:

1. Contracting path
2. Bottle-neck
3. Attention gating (This is the main contribution of the work)
4. Expanding path

3.3 Experiments and Results

Dataset and Implementation Details: The Attention U-Net model was evaluated on abdominal CT multi-label segmentation problem. The experiments were made on two different datasets: the first dataset consists of 150 abdominal 3D CT scans acquired from patients diagnosed with gastric cancer (CT-150)

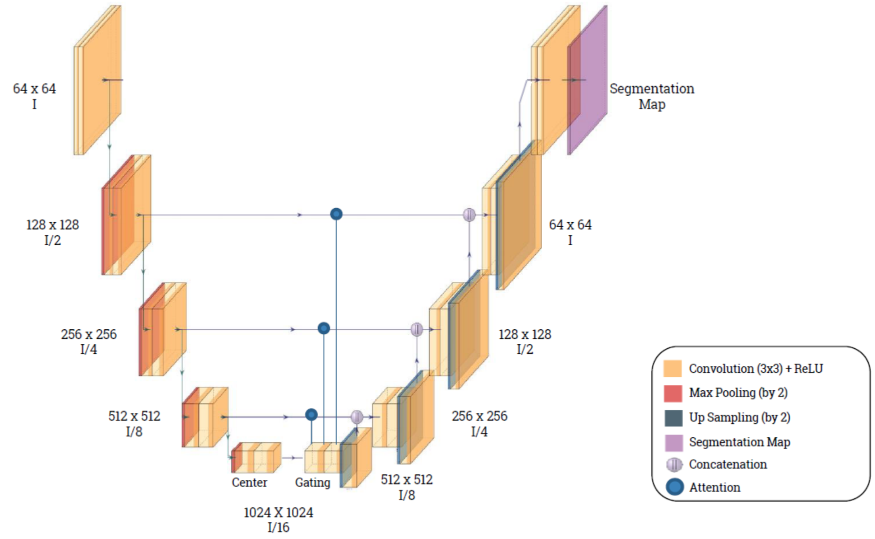


Fig. 4. Schematic of Attention U-Net architecture

and the second dataset (CT-82) consists of 82 contrast enhanced 3D CT scans with pancreas manual annotations. The models are trained using Sorensen-Dice loss [5], Adam optimizer, batch normalization and standard data-augmentation techniques including flips, transformations and random crops.

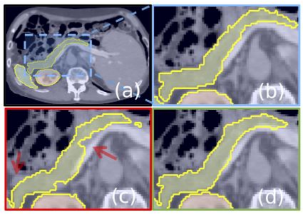


Fig. 5. (a) Ground truth Predictions: (b) U-Net (c) Attention U-Net (d) Missed

Segmentation Experiments and Comparisons: On multi-class abdominal CT segmentation dataset (CT-150), two different setups were made: one with 120 training images and 30 testing images and the other with vice versa of the first setup. The experiments performed using Attention U-Net was compared with the results of standard U-Net model. The Table 1), shows the Dice scores (DSC) and surface distances (S2S) of the two methods. The predictions are compared in Fig. 5). Conducting the experiments, it was observed that a little change in the capacity of the standard U-Net model improved the performance in terms of DSC by 2-3%. Also, the results from state-of-the-art CT pancreas segmentation models are summarized in Table 2.

Method (Train/Test Split)	U-Net (120/30)	Att U-Net (120/30)	U-Net (30/120)	Att U-Net (30/120)
Pancreas DSC	0.814±0.116	**0.840±0.087**	0.741±0.137	**0.767±0.132**
Pancreas Precision	0.848±0.110	0.849±0.098	0.789±0.176	**0.794±0.150**
Pancreas Recall	0.806±0.126	**0.841±0.092**	0.743±0.179	**0.762±0.145**
Pancreas S2S Dist (mm)	2.358±1.464	**1.920±1.284**	3.765±3.452	3.507±3.814
Spleen DSC	0.962±0.013	0.965±0.013	0.935±0.095	**0.943±0.092**
Kidney DSC	0.963±0.013	0.964±0.016	0.951±0.019	0.954±0.021
Number of Params	5.88 M	6.40 M	5.88 M	6.40 M
Inference Time	0.167 s	0.179 s	0.167 s	0.179 s

Table 1. Multi-class CT abdominal segmentation results obtained on the CT-150 dataset

Method	Dataset	Pancreas DSC	Train/Test	# Folds
Hierarchical 3D FCN [27]	CT-150	82.2 ± 10.2	Ext/150	-
Dense-Dilated FCN [6]	CT-82 & Synapse[3]	66.0 ± 10.0	63/9	5-CV
2D U-Net [8]	CT-82	75.7 ± 9.0	66/16	5-CV
Holistically Nested 2D FCN Stage-1[26]	CT-82	76.8 ± 11.1	62/20	4-CV
Holistically Nested 2D FCN Stage-2[26]	CT-82	81.2 ± 7.3	62/20	4-CV
2D FCN [4]	CT-82	80.3 ± 9.0	62/20	4-CV
2D FCN + Recurrent Network [4]	CT-82	82.3 ± 6.7	62/20	4-CV
Single Model 2D FCN [38]	CT-82	75.7 ± 10.5	62/20	4-CV
Multi-Model 2D FCN [38]	CT-82	82.2 ± 5.7	62/20	4-CV

Table 2. State-of-the-art CT pancreas segmentation methods that are based on single and multiple CNN models.

4 Discussion

4.1 Motivation

The main motivation of the paper is to improve localizing the organs of different structures and sizes. This will improve the performance of automated dense labeling on medical images. The motivation is valid, as in the domain of medical image analysis, labeling large amounts of medical images is difficult and involves human fatigue. Hence, this motivation brings up a special issue which lead the authors to propose the extended methodology on FCN architectures.

4.2 Research Questions

There are certain formulated research questions that the authors try to solve through the proposed methodology. They are as follows:

1. In multi-staged cascaded CNNs, can the repeated extraction of similar low-level features by all models be eliminated?
2. Can the involvement of large number of model parameters be eliminated?
3. Can the localization performance for labeling images be improved?

In order to solve these questions, the authors introduce the soft-attention mechanism to the fully convolutional networks.

4.3 Related Work

There are sufficient references provided for the state-of-the-art approaches on CT Pancreas Segmentation. It can be seen that valid and required references are provided for the main contribution, AGs, their usage in different domains and also different types of attention mechanisms available. Although these references seem convincing, most or all of the approaches are carried out in a 2D setting, but the proposed methodology is carried out in a 3D setting. Also, there is no supportive reference to why labeling large amounts of images is a tedious task and how will these accurate localization actually improve the automated medical image segmentation.

4.4 Methodology

The methodology is divided into three parts.

FCN: The basic idea of FCN and commonly used FCNs are portrayed. However, a clear view on how the FCNs are used with a 3D setting is not given. Also, it is not necessary to introduce the operations performed by a standard convolutional layer followed by ReLU activation. The attention mechanism between the input and gating layer is not demonstrated clearly in the Fig. 1. From the diagram, it seems like the attention gates are a result of the up-sampled layers and the skip connections, but it is not. Hence, after reviewing the methods and source code [15], a better image of the Attention U-Net architecture is shown in Fig. 4.

Attention Gates for Image Analysis: The parameter notations in the referenced figure (Fig. 2) and the text are not clearly correlated. It can be seen that in the text, the output of AGs is represented as $hatx^l_{i,c} = x^l_{i,c} * \alpha^l_i$, but the referenced figure has notations without i and c in the subscripts. Also, the representation of i and c in this context is not given. In the figure, $W_g and W_x$ are shown as 1x1x1 (3D), but in the text, they are shown as $W_x \in \mathbb{R}^{\mathrm{F_l x F_{int}}}$ and $W_g \in \mathbb{R}^{\mathrm{F_g x F_{int}}}$. $\mathbb{R}^{\mathrm{F_{int}}}$ could have been explained in detail.

Attention Gates in U-Net Model: The update rule for convolution parameters in different layers is formulated. However, the first paragraph provides the same information as discussed in the earlier parts of the work. U-Net model was proposed with a 2D setting, but the authors use with 3D setting. This transformation is not mentioned in the paper. Also, parts 2 and 3 in the methodology could have been merged.

4.5 Experiment and Results

The models are trained using Dice loss which is acceptable for this proposed method. Although the implementation and other details are provided, the hardware configuration and training time are not showed in the paper. The tables 1 and 2 could have been merged for better visualization and comparison of the results with and without higher capacity U-Net models. In table 4, state-of-the-art results are not compared with the proposed method and it can be seen from the table that most of the experiments were performed with 4-CV folds, 62/20(Train/Test) split and 2D setting. But the authors claimed that performance can be improved by carried out the experiments with 5-CV folds, 61/21(Train/Test) split and 3D setting. Hence, the results and analysis claimed with respect to the related work are questionable.

4.6 Conclusion

The answer to the research questions are provided briefly. The impact of their findings is not provided. However, the authors have discussed on their future work, ways to enhance the performance and other possible research directions related to this work.

4.7 Paper Format

The paper is grammatically correct and has no spelling errors. There are some basic suggestions related to the figures and tables. The titles used for figures and tables are not in a single line but in paragraphs. The continuous arrangement of tables (1, 2 and 3) and paragraphed titles make the complete page look messy. The method names used in Table 2 have similar splits (120/30) instead of 120/30 and 30/120. So, the results are not clear with respect to the splits. Also, the location of figures and tables are not ordered. For example, the figure 1 is shown in page 3, but the reference is made in page 4.

4.8 Paper Evaluation

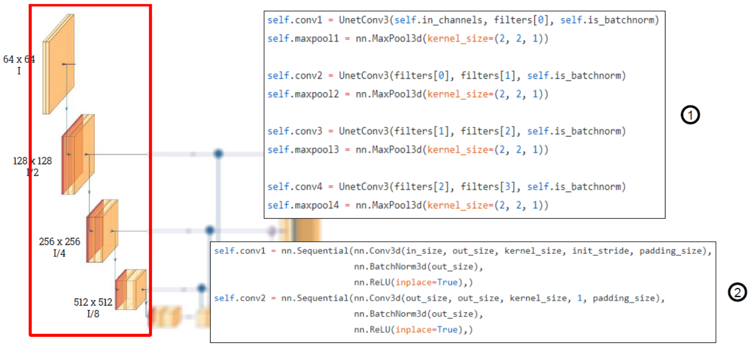
As discussed in the Introduction section, the paper has a total of 45 citations [14]. There are 3 versions of this paper available with edits and updates in each version. Initial version was submitted on 11 April 2018 and the latest revised version was submitted on 20 May 2018.

4.9 Code Review

To have a complete understanding on the proposed approach, the publicly available source code [14] is analysed and reviewed.

Contracting Path: Basic down sampling process

1. The 3D convolution layers and pooling layers are initialized.
2. The convolution followed by batch normalization and ReLU activation is done twice and then the feature maps are down sampled.



①

```
self.conv1 = UnetConv3(self.in_channels, filters[0], self.is_batchnorm)
self.maxpool1 = nn.MaxPool3d(kernel_size=(2, 2, 1))

self.conv2 = UnetConv3(filters[0], filters[1], self.is_batchnorm)
self.maxpool2 = nn.MaxPool3d(kernel_size=(2, 2, 1))

self.conv3 = UnetConv3(filters[1], filters[2], self.is_batchnorm)
self.maxpool3 = nn.MaxPool3d(kernel_size=(2, 2, 1))

self.conv4 = UnetConv3(filters[2], filters[3], self.is_batchnorm)
self.maxpool4 = nn.MaxPool3d(kernel_size=(2, 2, 1))
```

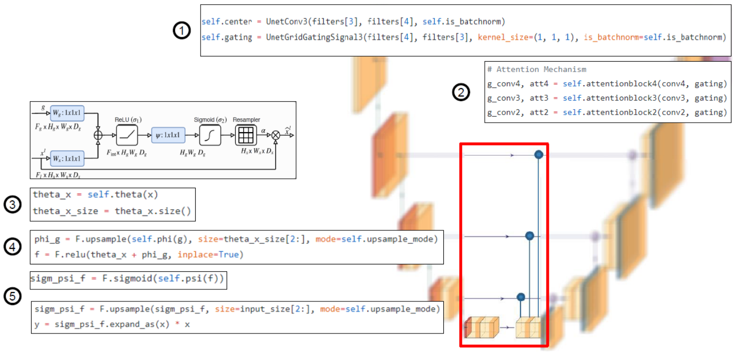
②

```
self.conv1 = nn.Sequential(nn.Conv3d(in_size, out_size, kernel_size, init_stride, padding_size),
                           nn.BatchNorm3d(out_size),
                           nn.ReLU(inplace=True),)
self.conv2 = nn.Sequential(nn.Conv3d(out_size, out_size, kernel_size, 1, padding_size),
                           nn.BatchNorm3d(out_size),
                           nn.ReLU(inplace=True),)
```

Fig. 6. Code Review - Contracting Path

Bottle-neck: Proposed attention mechanism is introduced in this part.

1. The center layer and gating layers are initialized.
2. Attention mechanism is carried out with different input convolutional layers and the gating layer.
3. Initializing gating parameters.
4. The operations of attention mechanism shown in formula (1) are carried out which results in the attention co-efficients.
5. Irrelevant regions are suppressed by multiplying the input layer element-wise with the output of AG.


①
②
③
④
⑤
g
$W_g: 1x1x1$
$F_g \times H_g \times W_g \times D_g$
ReLU (σ_1)
$\psi: 1x1x1$
Sigmoid (σ_2)
Resampler
α
$\hat{x}^l$
x^l
$W_x: 1x1x1$
$F_{int} \times H_g W_g D_g$
$H_g W_g D_g$
$H_x \times W_x \times D_x$
$F_l \times H_x \times W_x \times D_x$

①

```
self.center = UnetConv3(filters[3], filters[4], self.is_batchnorm)
self.gating = UnetGridGatingSignal3(filters[4], filters[3], kernel_size=(1, 1, 1), is_batchnorm=self.is_batchnorm)
```

②

```
# Attention Mechanism
g_conv4, att4 = self.attentionblock4(conv4, gating)
g_conv3, att3 = self.attentionblock3(conv3, gating)
g_conv2, att2 = self.attentionblock2(conv2, gating)
```

③

```
theta_x = self.theta(x)
theta_x_size = theta_x.size()
```

④

```
phi_g = F.upsample(self.phi(g), size=theta_x_size[2:], mode=self.upsample_mode)
f = F.relu(theta_x + phi_g, inplace=True)
```

```
sigm_psi_f = F.sigmoid(self.psi(f))
```

⑤

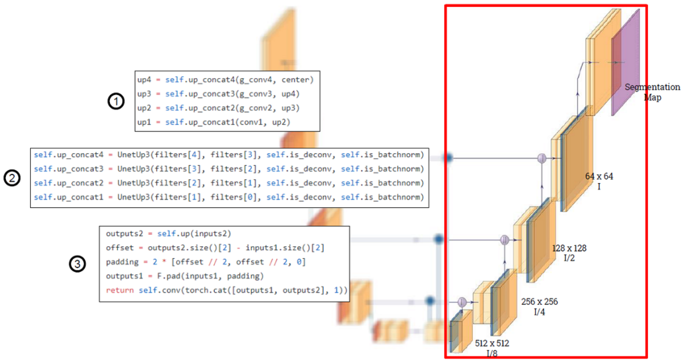
```
sigm_psi_f = F.upsample(sigm_psi_f, size=input_size[2:], mode=self.upsample_mode)
y = sigm_psi_f.expand_as(x) * x
```

Fig. 7. Code Review - Bottle-neck

Expanding Path: Upsampling and Concatenation.

1. The calling function for concatenating the attention gated skip-connection and the upsampled layer.

2. The layers are concatenated using this piece of code.
3. For upsampling, the layers are padded and then concatenated with the attentions.



```
up4 = self.up_concat4(g_conv4, center)
up3 = self.up_concat3(g_conv3, up4)
up2 = self.up_concat2(g_conv2, up3)
up1 = self.up_concat1(conv1, up2)
```

```
self.up_concat4 = UnetUp3(filters[4], filters[3], self.is_deconv, self.is_batchnorm)
self.up_concat3 = UnetUp3(filters[3], filters[2], self.is_deconv, self.is_batchnorm)
self.up_concat2 = UnetUp3(filters[2], filters[1], self.is_deconv, self.is_batchnorm)
self.up_concat1 = UnetUp3(filters[1], filters[0], self.is_deconv, self.is_batchnorm)
```

```
outputs2 = self.up(inputs2)
offset = outputs2.size()[2] - inputs1.size()[2]
padding = 2 * [offset // 2, offset // 2, 0]
outputs1 = F.pad(inputs1, padding)
return self.conv(torch.cat([outputs1, outputs2], 1))
```

Fig. 8. Code Review -Expanding Path

5 Conclusion

The Attention U-Net methodology, motivated by the automated identification and localization of organs for medical image analysis, shows a bit improvement in the performance compared to the state-of-the-art. The proposed method extends the convolutional networks with a soft-attention attention mechanism that learns to focus on the relevant regions in the skip-connected layers by suppressing the irrelevant regions. The impact of these output gates show better localization of the organs and the authors claim that the performance will be improved to a greater extent in their future work. The research direction related to the training behaviour of the AGs can benefit from transfer learning where pre-trained weights of U-Net model can be used to initialise the attention networks is also discussed in the paper. Finally, different techniques and approaches for improving the performance discussed by the authors provide motivation for the future works in the current research direction.

References

1. Bai, W., Sinclair, M., Tarroni, G., Oktay, O., Rajchl, M., Vaillant, G., Lee, A.M., Aung, N., Lukaschuk, E., Sanghvi, M.M., et al.: Human-level CMR image analysis with deep fully convolutional networks. In: arXiv preprint arXiv:1710.09289, (2017)
2. Jetley, S., Lord, N.A., Lee, N., Torr, P.: Learn to pay attention. In: International Conference on Learning Representations, (2018)
3. Liao, F., Liang, M., Li, Z., Hu, X., Song, S.: Evaluate the malignancy of pulmonary nodules using the 3D deep leaky noisy-or network. In: arXiv preprint arXiv:1711.08324, (2017)

4. Long, J., Shelhamer, E., Darrell, T.: Fully convolutional networks for semantic segmentation. In: IEEE CVPR. pp. 3431–3440, (2015)
5. Milletari, F., Navab, N., Ahmadi, S.A.: V-net: Fully convolutional neural networks for volumetric medical image segmentation. In: 3D Vision. pp. 565–571. IEEE, (2016)
6. Oda, M., Shimizu, N., Roth, H.R., Karasawa, K., Kitasaka, T., Misawa, K., Fujiwara, M., Rueckert, D., Mori, K.: 3D FCN feature driven regression forest-based pancreas localization and segmentation. In: DLMI, pp. 222–230. Springer, (2017)
7. Ronneberger, O., Fischer, P., Brox, T.: U-net: Convolutional networks for biomedical image segmentation. In: MICCAI. pp. 234–241. Springer, (2015)
8. Roth, H.R., Lu, L., Lay, N., Harrison, A.P., Farag, A., Sohn, A., Summers, R.M.: Spatial aggregation of holistically-nested convolutional neural networks for automated pancreas localization and segmentation. In: Medical Image Analysis 45, 94 – 107, (2018)
9. Roth, H.R., Oda, H., Hayashi, Y., Oda, M., Shimizu, N., Fujiwara, M., Misawa, K., Mori, K.: Hierarchical 3D fully convolutional networks for multi-organ segmentation. In: arXiv preprint arXiv:1704.06382, (2017)
10. Saito, A., Nawano, S., Shimizu, A.: Joint optimization of segmentation and shape prior from level-set-based statistical shape model, and its application to the automated segmentation of abdominal organs. In: Medical image analysis 28, 46–65, (2016)
11. Wang, F., Jiang, M., Qian, C., Yang, S., Li, C., Zhang, H., Wang, X., Tang, X.: Residual attention network for image classification. In: IEEE CVPR. pp. 3156–3164, (2017)
12. Wolz, R., Chu, C., Misawa, K., Fujiwara, M., Mori, K., Rueckert, D.: Automated abdominal multi-organ segmentation with subject-specific atlas generation. IEEE TMI 32(9) (2013)
13. Zografos, V., Valentinitsch, A., Rempfler, M., Tombari, F., Menze, B.: Hierarchical multi-organ segmentation without registration in 3D abdominal CT images. In: International MICCAI Workshop on Medical Computer Vision. pp. 37–46. Springer, (2015)
14. Citation counts taken from https://scholar.google.com/
15. Attention U-Net: https://github.com/ozan-oktay/Attention-Gated-Networks
16. International Conference on Medical Imaging with Deep Learning: https://www.midl.io/aims-and-scope.html