OPEN SET SEMANTIC SEGMENTATION WITH STATISTICAL TEST AND ADAPTIVE THRESHOLD

Zhiying Cui[*], *Longshi Wu*[*], *Ruixuan Wang*[(✉)]

cuizhy6@mail2.sysu.edu.cn, wulsh26@mail2.sysu.edu.cn,
wangruix5@mail.sysu.edu.cn.

ABSTRACT

Semantic segmentation in the open world is prerequisite when deploying a well-trained segmentation model in real scenarios, where objects of unseen classes during model training may often appear in future new images to be segmented by the model. However, such open set semantic segmentation task has been rarely explored before. In this study, making use of the large number of pixel-level prediction uncertainties for each image, we proposed applying the non-parametric statistical test to detect whether objects of unseen classes appear in a new image, and an adaptive threshold method to automatically segment each pixel into either one of the known classes or the unknown class. Experiments on the natural image dataset showed that the proposed method significantly outperforms multiple strong baseline methods.

Index Terms— open set, semantic segmentation, adaptive threshold, statistical test.

1. INTRODUCTION

Semantic segmentation aims to segment an image into multiple semantic regions, often by classifying each image pixel into one of the predefined classes based on deep learning models, such as UNet [1], DeepLab [2, 3], and attention Network [4, 5]. In most semantic segmentation studies, people assume that the categories appearing in test images are the same as those in the training images, i.e., the segmentation is based on a close set of categories. However, the world is open in real scenarios, i.e., objects or things of novel categories could appear in test images but do not appear in training images. This often happens in real applications, such as autonomous driving and medical image diagnosis, where it is almost impossible to collect and annotate all categories of object or disease regions for the training of semantic segmentation models. In this case, it would be ideal if the segmentation model can not only segment image regions of the previously seen categories, but also tell users the region of any unseen category (if existing in a new image) does not belong to those categories learned during model training. This task can be called open set semantic segmentation (Figure 1).

[*] The authors contribute equally to this paper.

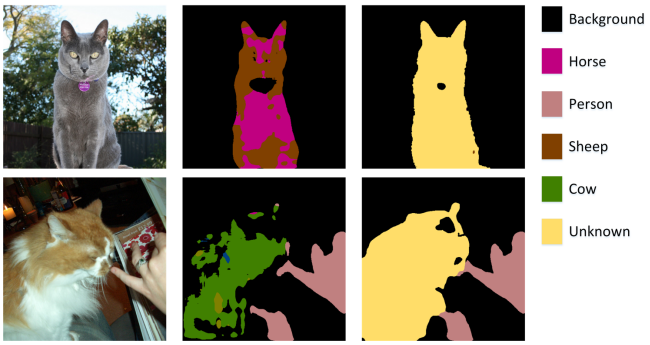

Fig. 1. Examples of open set semantic segmentation. Left column: input images in which regions of *cat* belongs to the unknown class; Middle: segmentation results in close set semantic segmentation where any unknown region was forced to be segmented into known class(es); Right: open set semantic segmentation result based on the proposed method.

While open set semantic segmentation has been rarely explored, the open set recognition has been extensively studied recently, where the objective is to classify any new image into one of the previously seen (i.e., *known*) classes or the unseen (i.e., *unknown*) class. One approach explores the characteristics of the classifier outputs particularly for images of the unknown class, and finds that the maximum output probability (over all possible known classes) is often relatively small if the input image is from the unknown class [6, 7, 8], while the maximum output probability is more likely close to 1.0 for images of known classes. In other words, the output prediction is relatively uncertain for images of the unknown class. With this observation, people can determine whether a new image is from the unknown class by comparing the maximum output probability with a pre-defined threshold or by comparing the output prediction uncertainty (often represented by the entropy of the output probability distribution) with a predefined threshold [7]. Another approach is to train generative models such that the distribution of each known class can be explicitly or implicitly represented [9]. Then a new image can be detected as the unknown class if the image is far from the distribution of each known class, where the distance can

978-1-7281-1331-9/20/$31.00

be represented by Mahalanobis distance between the new image and each class center [10] in the image feature space, or by the reconstruction error between the original image and the reconstructed image as used in the auto-encoder models [11, 12, 13]. These two approaches have the difficulty of choosing an appropriate threshold to determine whether a new image is from a known class or from the unknown class [14]. In contrast, the third approach explicitly construct pseudo images for the unknown class by generating adversarial examples based on the images of known classes [10], and then use the images of known classes and the pseudo images of the unknown class to directly train a classifier which can classify both known classes and the unknown classes.

Although semantic segmentation is often regarded as dense classification of image pixels, at least some of the approaches for the open set recognition may not be easily modified to solve the open set semantic segmentation task. For example, the adversarial example approach might not be easily transferred to open set semantic segmentation because it is not clear how to generate adversarial examples for every image pixels. The Mahalanobis distance based approach [10] also might not be easily adapted because the distribution of pixel-level features within each class may be multi-mode and therefore does not satisfy the underlying assumption of Gaussian distribution. On the other hand, some other approaches for open set recognition seems feasible for open set semantic segmentation. In this study, the maximum output probability method and the entropy-based uncertainty method [7] were applied as baselines for open set semantic segmentation.

To solve the difficulty of determining the threshold existed in most approaches, we make use of the characteristics of open set semantic segmentation, i.e., multiple output predictions can be collected for each image, with each prediction corresponding to one image pixel, and propose an adaptive method to automatically determine the threshold for each image segmentation. The effectiveness of the adaptive threshold was confirmed with the entropy-based uncertainty method on multiple open set semantic segmentation tasks. Another contribution of this study is the application of the non-parametric statistical test to the determination of existence of the unknown category in any new image. Experiments show that correctly determining the existence of the unknown category (as the first step) can help effectively reduce the false segmentation of some known pixels into the unknown category. Last but not least, we also summarized the challenges existed in the rarely explored open set semantic segmentation tasks for future study.

2. OPEN SET SEMANTIC SEGMENTATION

Open set semantic segmentation aims to classify each image pixel into either one of the existing (i.e., *known*) classes or the *unknown* class, where the unknown class is often the combination of many other classes not appearing in the training dataset. The segmentation model need to be trained only based on a set of training images without the unknown class therein. Following the finding in open set classification that the classifier output prediction is often relatively uncertain (i.e., the classifier's output probability distribution is more spread across multiple classes, often represented by the *entropy* of probability distribution) for images of the unknown class, here we investigate the utility of prediction uncertainty for each image pixel in the open set semantic segmentation task, where the uncertainty u_i for the i-th pixel of an image can be represented by the entropy of the output prediction probability $\mathbf{p}_i = (p_{i,1}, p_{i,2}, \ldots, p_{i,C})^{\mathrm{T}}$, i.e.,

$$u_i \quad = \quad -\sum_{c=1}^{C} p_{i,c} \log p_{i,c} \,. \tag{1}$$

One reasonable presumption is that the predictions for pixels of the unknown class would be more uncertain than those for pixels of known classes. However, given any new image without the unknown class therein, relatively high prediction uncertainty has been observed particularly around the boundary of objects within the image. Therefore, given the probability prediction for each pixel, it is required to reliably determine whether there really exist pixels or regions of the unknown class, and if existing, where the pixels or regions are. Here we propose applying a statistical test method to determine the existence of the unknown class and an adaptive threshold method to find pixels of the unknown class for any new image.

2.1. Determination of unknown-class existence

After the segmentation model is trained based on a set of training images without the unknown class inside, correctly determining the existence of unknown class in a new image would clearly improve the performance of open set semantic segmentation, particularly in reducing the false classification of pixels of the known classes into the unknown class. The key challenge is to determine whether the pixels with higher prediction uncertainties are from the unknown class or from part (e.g., boundary) of objects of known classes. Assuming that the number of pixels with relatively high prediction uncertainties is very small in images without the unknown class (Figure 3, first two), while there are more pixels with relatively higher prediction uncertainty in images with certain unknown regions inside (Figure 3, last two), we propose applying the non-parametric statistical Mann-Whitney U test to determine the existence of the unknown class in any new image. The non-parametric U test instead of the parametric t-test is adopted because the distribution of prediction uncertainties is often long-tailed (Figure 3) and therefore the parametric test is not feasible.

The procedure of the U test is as follows. First, a small proportion α ($\alpha = 0.5\%$ used in experiments) of pixels with

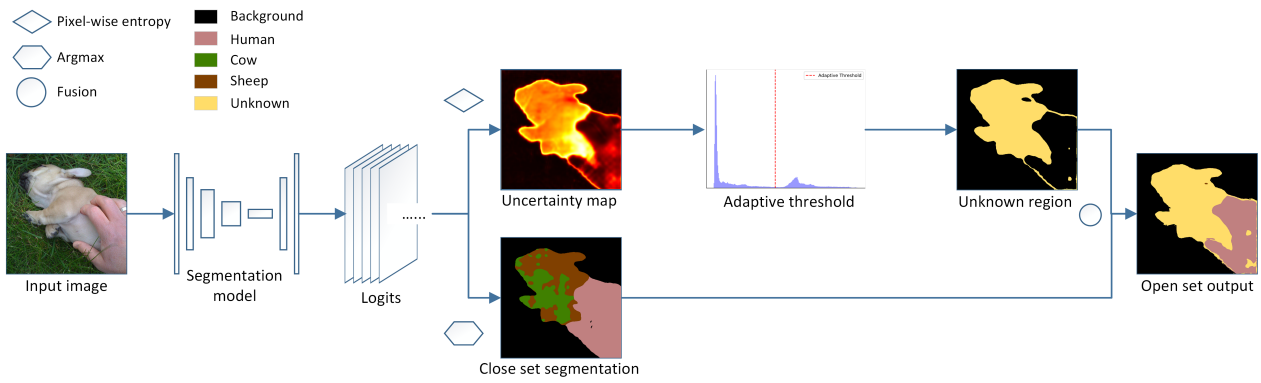

Fig. 2. The pipeline of open set semantic segmentation. The region of the unknown class is segmented based on the adaptive threshold (upper path) and then fused with the initial segmentation result (lower path) to generate the final open set segmentation result. Dog is the unknown object which never appears in the training set, while *Human* is one known class.

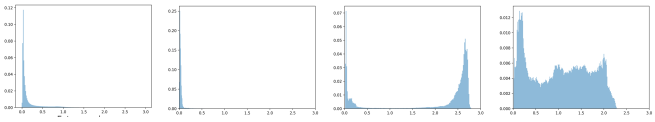

Fig. 3. Distributions (histograms) of prediction uncertainties over pixels of all validation images (first), of one test image without the unknown class therein (second), and of two other test images including regions of the unknown class (last two).

the highest prediction uncertainties are first collected over a set of validation images in which there is no unknown class, and then these pixels are sub-sampled to typically represent the small proportion of pixels with high prediction uncertainty for an virtual average validation image. For any new image, a similar small proportion of pixels with the highest prediction uncertainties are also collected. Finally, the Mann-Whitney U test is used to test whether the prediction uncertainties of these pixels from the new image are significantly higher than the prediction uncertainties of the previously sub-sampled pixels. It is expected that there would be no such one-tail significant difference if the new image does not contain the unknown class, while there would be for a new image containing unknown regions (due to higher prediction uncertainties from the unknown regions). Note that such statistical test method is not suitable for open set classification tasks because there is just one prediction uncertainty value for one image in the classification tasks. In this sense, we consider the application the statistical Mann-Whitney U test to the open set semantic segmentation to be novel.

2.2. Adaptive threshold for segmentation of unknown-class regions

Once a new image has been determined to contain region(s) of the unknown class, the next is to automatically find pixels belonging to the unknown-class region. Since in general the unknown-class region occupies just part of the image, the prediction uncertainties of the known-class pixels would be generally much lower. As a result, the distribution (here approximated by a histogram) of the prediction uncertainties over all pixels of the image would have at least two modes (peaks), with the mode of smaller uncertainties mainly contributed by the pixels of known classes and the mode of larger uncertainties contributed potentially by the pixels of the unknown class (Figure 3, last two histograms). Such multi-mode property provides us with a simple but effective method to identify the pixels of the unknown-class, i.e., find one local minimum (a valley between peaks) of the uncertainty distribution and then all those pixels whose prediction uncertainties are higher than the minimum would be considered as the unknown class. It is clear that the local minimum in general would be different for different images (Figure 3, last two histograms), thus calling local minimum as *adaptive threshold*. In contrast, a fixed threshold (minimum) could be used for all images, but such fixed threshold in general performs worse than the proposed adaptive threshold (see Experiment section 3.2).

One question may arise if there are two or more local minima (i.e., three or more peaks) in the uncertainty distribution, i.e., which local minimum shall we choose as the adaptive threshold for the image? One may choose the largest (rightmost in the histogram), the smallest (leftmost in the histogram), or the global minimum (the entropy with smallest frequency in the histogram) between the peaks. We argue that the choice strategy of local minimum could be determined by users in real applications. For example, in autonomous driving or intelligent medical diagnosis, missing segmentation of unknown-class object or regions could cause serious consequences, therefore users may choose the leftmost local minimum as the adaptive threshold to avoid any potential missing of unknown objects or disease regions. In this study, we ex-

Table 1. Dataset construction. Training and validation sets do no include any object of the unknown class.

Unknown class	Train	Valid	Test	
			known	uknown
cat & *dog*	1215	634	572	492
car & *bus*	1307	936	309	361
bike & *motor*	1344	963	305	301

perimentally show that either way of adaptive threshold performs better than strong baseline methods.

3. EXPERIMENT

3.1. Experimental setup

Since open set semantic segmentation has been rarely explored, there is no public dataset for such a task. Here we adapted Pascal VOC 2012 to the condition of open set semantic segmentation. More specifically, from all the 21 classes in the dataset, two classes (e..g, *cat* and *dog*) were pre-selected to form the *unknown* class and all the images containing either of the two pre-selected classes were left out as part of the test set. For the other images containing the remaining 19 classes, around 1200-1300 images were randomly selected as the training set, and 600-1000 other images were used as the validation set (which does not include the unknown class as well). Such data split were performed three times, with the unknown class being different each time (See Table 1 for details). Segmentation models (e.g., DeepLabV3) were then trained and evaluated with each of the created open set segmeantation datasets.

During training of each segmentation model, SGD optimizer was used with the initial learning rate 0.007, the momentum value 0.9, and the coefficient of weight decay 0.001. The learning rate was updated with the Poly strategy. In evaluation, the significance level was set 0.05 during the Mann-Whitney U test and the adaptive threshold was determined by the rightmost, the leftmost, and the global minimum of the entropy histogram respectively for each image. Since the performance is similar between these threshold choice strategies, only the performance with the global minimum based adaptive threshold was reported. The Intersection-over-Union (IoU) was used as the metric to measure the performance of the proposed method for segmentation of the unknown class, and the mean IoU was used to measure the performance on the segmentation of the known classes. For each test, five runs were performed and the average and standard deviation of the IoU or the mean IoU were reported.

3.2. Comparison with baseline methods

The proposed method was firstly evaluated on the Pascal dataset by comparing with a few baseline methods, with DeepLabV3+ as the model backbone. The maximum output based method ODIN [15] and the entropy-based uncertainty method originally for open set recognition were used as two baselines considering their feasibility for the open set semantic segmentation. Note that these two baselines do not provide the way to choose an appropriate threshold for segmentation. Here the optimal threshold with the highest mean IoU over all classes was greedily searched for each of the two baselines. In addition, to evaluate the effectiveness of the proposed adaptive threshold, a fixed threshold across all test images was used to replace the adaptive threshold and then combined with the first step (i.e., determining existence of the unknown class) of the proposed method as another baseline ('Utest+Fixed' in Table 2). The fixed threshold was estimated on the validation set such that the proportion of pixels whose maximum output probabilities are larger than the threshold is equal to a preset level β. β was respectively set to $0.005, 0.01, 0.02, 0.05, 0.1, 0.15$ and only the one (0.05) with the best performance was reported. To evaluate the effectiveness of the first step of the proposed method, the performance of the adaptive threshold without the first step was also reported ('Adaptive' in Table 2).

From Table 2, we can see that the proposed method (last row) significantly outperforms not only the baselines adapted from open set recognition (also see Figure 4), but also the baselines 'Utest+Fixed' and 'Adaptive'. In particular, the U test can also help improve the segmentation performance when combined with the fixed threshold method, probably because U test can well tell which images did not contain the unknown class and therefore avoided the error of segmenting pixels into the unknown class in these images. Compared to the fixed threshold, adaptive threshold together with the U test further improved the segmentation performance particularly on the regions of the unknown class, confirming that the adaptive threshold can find a better image-specific threshold for segmentation of the unknown class. All these suggest that both the Mann-Whitney U test and the adaptive threshold in the proposed method are helpful in improving the performance of open set semantic segmentation.

3.3. Evaluation with different backbone models

The proposed method was also evaluated with different deep segmentation model architectures, including the DeepLabV3 [3] used above, the well-known U-Net [1] and the EMANet [5]. DeepLabV3 is based on dilated convolution with multiple scale poolings; U-Net makes use of skip connections between encoder layers and corresponding decoder layers, and EMANet uses attention mechanism during segmentation. Therefore, these three models represent three different types of architectures for semantic segmentation. Initial results (not shown due to limited space) tell us that the proposed method is always better than the baselines on all the three model architectures, supporting that the proposed

Table 2. Performance of the proposed method and baselines. Each *unknown* column: IoU for the unknown class; each *known* column: mean IoU over all the known classes; each *all* column: mean IoU over the unknown and all the known classes. The standard deviation of the (mean) IoU is around 0.004-0.006, and the backbone is DeepLabV3+ for all methods.

Method	*cat & dog*			*car & bus*			*bike & motor*		
	unknown	*known*	*all*	*unknown*	*known*	*all*	*unknown*	*known*	*all*
Entropy	0.5131	0.6317	0.6258	0.3972	0.3949	0.395	0.3528	0.6842	0.6676
ODIN	0.5293	0.6339	0.6287	0.3852	0.3268	0.3297	0.3414	0.65	0.6346
Utest+Fixed	0.415	0.6099	0.6002	0.1892	0.6338	0.6116	0.1772	0.7265	0.6991
Adaptive (No UTest)	0.487	0.5685	0.5644	0.3289	0.5815	0.5689	0.3805	0.7149	0.6982
Utest+Adaptive (Ours)	0.5593	0.6493	**0.6448**	0.2933	0.6301	**0.6133**	0.3356	0.7296	**0.7099**

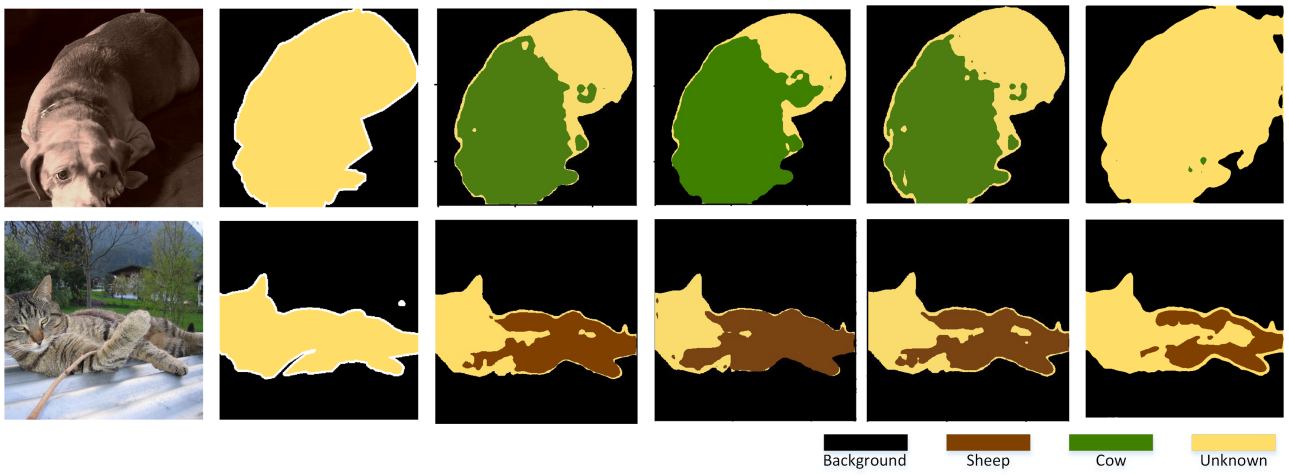

Fig. 4. Examples of open set segmentation results with various methods. From left to right: original input images, ground-truth segmentation maps, segmentation by entropy, ODIN, fixed threshold combined with the U test, and the proposed method.

method is robust to model architectures.

3.4. Challenges in open set semantic segmentation

We investigated the detailed open set segmentation results and observed a few challenging conditions in the open set segmentation task.

- Unseen objects in known classes (Figure 5, first row): it is almost impossible to collect and annotate a large number of image regions for each known class, and as a result, the segmentation model may be over-trained and cannot correctly segment some image regions of known classes. This is particularly true for the (known) *background* class, in which objects not appearing in the training set are often considered as part of the background region in the test images. In this case, it seems reasonable if the open set segmentation model considers such object regions as the unknown class. To more objectively evaluate various methods, it is necessary to create a public dataset particularly for the open set semantic segmentation task, carefully defining each known and unknown class.

- Unknown-class objects similar to known classes (Figure 5, second row): it is possible that certain objects of the unknown class are similar to the some objects of known classes. For example, the *cat* in the example (second row) is very similar to the *sheep* class, and therefore the segmentation model segment the region of the *cat* into the *sheep* class with high confidence. In this case, it is very challenging for the segmentation model to more precisely learn to separate not only among the known classes, but also between the known and (unseen) unknown class.

- Small unknown objects versus boundary of known objects (Figure 5, third row): another challenging condition is to accurately segment the small objects of the unknown class. Since boundaries of known objects are often causing higher prediction uncertainty, it is difficult to tell whether the small number of pixels with higher prediction uncertainty are really from an small unknown object or from the boundaries of known objects. Considering higher-level information (such as shape) during open segmentation may help disambiguate this confusion.

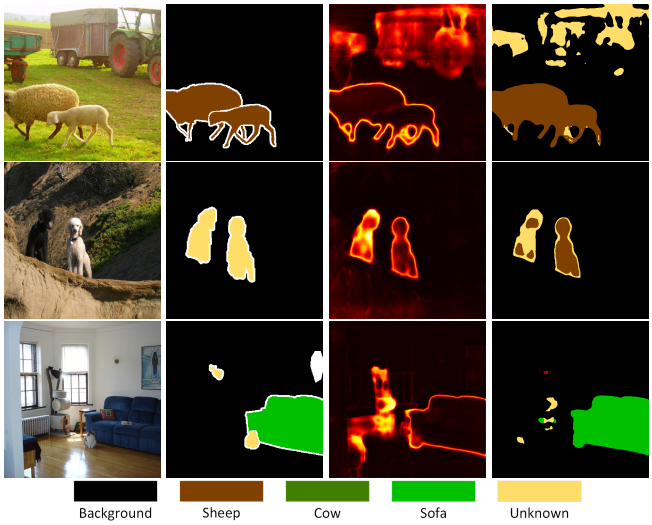

Fig. 5. Challenging conditions in open set semantic segmentation. First row: unseen objects of know classes; Second row: unknown object similar to known classes; Third row: small unknown object. First column: original test images; Second column: ground-truth segmentation maps; Third column: prediction uncertainty (entropy) maps; Last column: open set semantic segmentation with the proposed method (similar results obtained with other baseline methods).

An ideal open set semantic segmentation model need to effectively handle all the above challenging conditions, and much future study is required to solve this relatively new task.

4. CONCLUSIONS

Open set semantic segmentation is a rarely explored topic. In this study, making use of the large number of pixel-level prediction uncertainties for each image, we proposed the application of the non-parametric statistical test to the determination of unknown-class existence and an adaptive threshold method to automatically segment each pixel into either one of the known classes or the unknown class. Experiments on the natural image dataset showed that the proposed method significantly outperforms existing methods originally for open set recognition tasks. Some challenging conditions were also identified for future work.

Acknowledgement: This work is supported in part by the National Key Research and Development Program (grant No.2018YFC1315402), the Guangdong Key Research and Development Program (grant No.2019B020228001), the National Natural Science Foundation of China (grant No.U1811461), and the Guangzhou Science and Technology Program (grant No.201904010260).

5. REFERENCES

[1] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," *CoRR*, vol. abs/1505.04597, 2015.

[2] L. Chen, G. Papandreou, F. Schroff, and H. Adam, "Rethinking atrous convolution for semantic image segmentation," *CoRR*, vol. abs/1706.05587, 2017.

[3] L. Chen, Y. Zhu, G. Papandreou, F. Schroff, and H. Adam, "Encoder-decoder with atrous separable convolution for semantic image segmentation," *CoRR*, vol. abs/1802.02611, 2018.

[4] J. Fu, J. Liu, H. Tian, Z. Fang, and H. Lu, "Dual attention network for scene segmentation," *CoRR*, vol. abs/1809.02983, 2018.

[5] X. Li, Z. Zhong, J. Wu, Y. Yang, Z. Lin, and H. Liu, "Expectation-maximization attention networks for semantic segmentation," in *CVPR*, pp. 9167–9176, 2019.

[6] D. Hendrycks and K. Gimpel, "A baseline for detecting misclassified and out-of-distribution examples in neural networks," *CoRR*, vol. abs/1610.02136, 2016.

[7] D. Hendrycks and K. Gimpel, "A baseline for detecting misclassified and out-of-distribution examples in neural networks," *CoRR*, vol. abs/1610.02136, 2016.

[8] A. Kendall, V. Badrinarayanan, and R. Cipolla, "Bayesian segnet: Model uncertainty in deep convolutional encoder-decoder architectures for scene understanding," *CoRR*, vol. abs/1511.02680, 2015.

[9] T. DeVries and G. W. Taylor, "Learning confidence for out-of-distribution detection in neural networks," *arXiv preprint arXiv:1802.04865*, 2018.

[10] K. Lee, K. Lee, H. Lee, and J. Shin, "A simple unified framework for detecting out-of-distribution samples and adversarial attacks," in *NeurIPS*, pp. 7167–7177, 2018.

[11] D. Gong, L. Liu, V. Le, B. Saha, M. R. Mansour, S. Venkatesh, and A. van den Hengel, "Memorizing normality to detect anomaly: Memory-augmented deep autoencoder for unsupervised anomaly detection," *CoRR*, vol. abs/1904.02639, 2019.

[12] D. Abati, A. Porrello, S. Calderara, and R. Cucchiara, "Latent space autoregression for novelty detection," in *CVPR*, 2019.

[13] T. Schlegl, P. Seeböck, S. M. Waldstein, G. Langs, and U. Schmidt-Erfurth, "f-anogan: Fast unsupervised anomaly detection with generative adversarial networks," *MIA*, vol. 54, pp. 30 – 44, 2019.

[14] H. Choi, E. Jang, and A. A. Alemi, "Waic, but why? generative ensembles for robust anomaly detection," *arXiv preprint arXiv:1810.01392*, 2018.

[15] S. Liang, Y. Li, and R. Srikant, "Enhancing the reliability of out-of-distribution image detection in neural networks," in *ICLR*, 2018.