Active Contour Segmentation based on Histograms and Dictionary Learning for Videocapsule Image Analysis

Gaetan Raynaud, Camille Simon-Chane[a], Pierre Jacob and Aymeric Histace[b]
ETIS UMR 8051, UPS, UCP, ENSEA, CNRS, 6 av. du Ponceau, 95014, Cergy, France

Keywords: Active Contour, Bag of Words, Small Bowel Videocapsule.

Abstract: This article deals with statistical region-based active contour segmentation using histograms and dictionary learning. Following previous publication, the active contour segmentation using optimization alpha-divergence family, leads to satisfying results. The method of the segmentation is based on histograms of the luminance of the pixels. To improve this method and to allow it to adapt to more types of images, we propose to replace luminance histograms with histograms of features using a bag of features model. This approach will be able to overcome the limitations of the luminance and give a better representation of the image. We will present the approach to create the new representation of the image, first with associated histograms to show its potential, using a local approach based on dictionary learning to compute the probability map of each pixel of the image to belong to the targeted object. In a second step using histograms based on bag of features for the representation of the image. We present experiments for the two methods on images extracted from small bowel videocapsule acquisitions and for two types of targeted objects (angiodysplasia and ulcer). We show that by replacing the luminance representation by a more complex one, we reach better performances for the segmentation of the targeted objects.

1 INTRODUCTION

The purpose of segmentation is to extract homogeneous image regions representing objects. Active contour segmentation methods, first introduced by (Kass et al., 1988), consist of an iterative process that applies a velocity to a curve to fit the boundaries of the targeted object. The forces are applied from the inside and the outside of the curve. The most common approach is to use energies related to the segmentation problem. The existing methods can be divided into two categories: edge-based models (Kass *et al.*, 1988), Casselles *et al.*, 1997), and region-based models (Chan and Vese, 2001). One of the most popular edge-based models is the Geodesic active contours (Casselles *et al.*, 1997) which minimize the curve's length based on a function of the gradient of the image. The region-based methods like (Chan and Vese, 2001) were introduced in order to overcome the limitations of the edge-based methods. The region-based methods use statistical descriptors like mean or variance, from the inside and the outside of the region, which allows sturdiness to noise and better performance with weak edges than the edge-based methods.

Another approach, inspired by the region-based approach, is the statistical-region based active contour. This last approach intents to improve the region-based descriptors, like mean or variance of pixels, that fail to segment regions in image that cannot be easily discriminated by their first order statistics. (Aubert *et al.*, 2003) and (Lecellier *et al.*, 2009) proposes to use probability density function (PDF), describing both the inner (Ω_{in}) and outer (Ω_{out}) regions, to evolve the curve. The strategy consists in a minimization of the distance between the PDFs of inner and outer regions and two reference PDFs describing the targeted object and the background of the image. The usual way to do so consists in deriving the related steering partial differential equation (PDE) of the energy of the distance between the PDFs.

The key point of the method is to choose the distance function between two PDFs. Some of the most used distance are the Kullback-Leibler divergence (KL), the Hellinger distance or the χ^2 divergence. In this paper, we will use alpha-divergences as proposed

[a] https://orcid.org/0000-0002-4833-6190
[b] https://orcid.org/0000-0002-3029-4412

Raynaud, G., Simon-Chane, C., Jacob, P. and Histace, A.
Active Contour Segmentation based on Histograms and Dictionary Learning for Videocapsule Image Analysis.
DOI: 10.5220/0007694706090615
In *Proceedings of the 14th International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications (VISIGRAPP 2019)*, pages 609-615
ISBN: 978-989-758-354-4

by (Meziou *et al.*, 2011). The alpha-divergence between two PDFs p_1 and p_2 is given by Eq. (1):

$$\mathrm{D}_{\alpha}(\mathrm{p}_1 \| \mathrm{p}_2, \Omega) = \int_{\Omega} \varphi_{\alpha}(\mathrm{p}_1, \mathrm{p}_2, \lambda) \mathrm{d}\lambda \quad (1)$$

with $\varphi_{\alpha}(p_1, p_2, \lambda) =$

$$\frac{\alpha \mathrm{p}_1 + (1 - \alpha)\mathrm{p}_2 - {\mathrm{p}_1}^{\alpha} \, {\mathrm{p}_2}^{1-\alpha}}{\alpha(1-\alpha)} \quad (2)$$

In the case of the alpha-divergence between the two PDFs of the inner and outer regions, the corresponding energy J is given by Eq. (3):

$$\mathrm{J}(\Gamma, \Omega_{in}, \Omega_{out}) = \int_{\Re} \varphi_{\alpha}(\mathrm{p}_{in}, \mathrm{p}_{ref\_in}, \lambda) \mathrm{d}\lambda + \int_{\Re} \varphi_{\alpha}(\mathrm{p}_{out}, \mathrm{p}_{ref\_out}, \lambda) \mathrm{d}\lambda + \beta \int_{\Gamma} \mathrm{ds} \quad (3)$$

where Γ is the boundary between Ω_{in} and Ω_{out}, φ_{α} is the alpha-divergence between the two PDFs, the last term is the regularization of the contour with β being a positive constant. The PDFs are calculated using Parzen window approach.

Calculating the Euler derivative of (3), leads to evolution equation (4) of Γ:

$$\frac{\partial \Gamma}{\partial \mathrm{t}} = \Big[D(\Omega_{in}) - D(\Omega_{out}) + C(\Omega_{out}) - C(\Omega_{in}) + \frac{1}{|\Omega_{in}|} \partial_1 \varphi_{in}(\mathrm{p}_{in}, \mathrm{p}_{ref\_in}, \lambda) * g_{\sigma}(\mathbf{I}(\mathbf{x})) - \frac{1}{|\Omega_{out}|} \partial_1 \varphi_{out}(\mathrm{p}_{out}, \mathrm{p}_{ref\_out}, \lambda) * g_{\sigma}(\mathbf{I}(\mathbf{x})) + \beta\Big] \mathbf{N} \quad (4)$$

With

$$\begin{cases} D(\Omega_i) = \int_{\Omega_i} \varphi(\mathrm{p}_i, \mathrm{p}_{ref\_i}, \lambda) \mathrm{d}\lambda \\ C(\Omega_i) = \int_{\Omega_i} \partial_1 \varphi(\mathrm{p}_i, \mathrm{p}_{ref\_i}, \lambda) \mathrm{p}_i \, \mathrm{d}\lambda \\ \quad i = \{in, out\} \end{cases} \quad (5)$$

In (4) and (5), $\partial_1 \varphi$ denotes the first order derivative of φ function with respect to the estimated PDFs, g_{σ} is the Gaussian kernel (with standard-deviation σ) used in estimation of the PDFs of the two regions, **I(x)** is the statistical function representing the segmented image at the pixel x, and **N** the inward local normal vector of the moving curve Γ.

In this article, the optimization of the φ function is not the main purpose; a method to do so is presented in (Meziou *et al.*, 2014). As a consequence, we use the α parameter as a constant of $\frac{1}{2}$, leading the alpha-divergence to correspond to:

$$D_{\frac{1}{2}}(\Omega) = 2D_{Hellinger}(\Omega)$$

In fact, we propose a method to improve the statistical-region based active contour used in (Meziou *et al.*, 2014) in which the histogram used for the statistical representation, only takes into account the statistical luminance distribution of the data in the image. This paper proposes to replace the luminance by more complex representation and to use bags of features histograms as statistical representation.

The paper is organized as follow: In section 2, we begin by introducing the bags of features method, how to create them, and show some results. Then section 3 presents how we can use this approach to improve the statistical-region based method and the results on medicals images related to gastrointestinal image extracted from small-bowel videocapsules acquisition. Finally, section 4 concludes the paper.

2 LOCAL PROBABILITY USING DICTIONARY LEARNING

2.1 Method

In this section, we want to test if a bag-of-features approach can be a promising idea and if it can be fruitfully used in the context of histogram-based approach.

The alpha-divergence method previously introduced is based on histograms analysis to perform the segmentation. however as mentioned before the luminance histogram can be not discriminating enough for complex segmentation task where usual hypothesis (Gaussian distribution) are not fulfilled. Thus, it could be interesting to look for another way to represent the statistical information of the image; the closest method of the classic luminance histograms being the bag of words, we investigated this approach. To represent an image as a bag of words model, we need first to define the *words* that will form the *dictionary*, that is to say, the patterns that will embed the most representative statistical properties of the objects to segment in a further step. A good descriptor for those words is the Scale-invariant feature transform (SIFT) descriptor for instance. These descriptors will allow to get the interest (or most saillant) points inside the image. Considering those extracted interest points, we can use the SIFT descriptor, to describe interest areas around them that will be used for the bag of features dictionary.

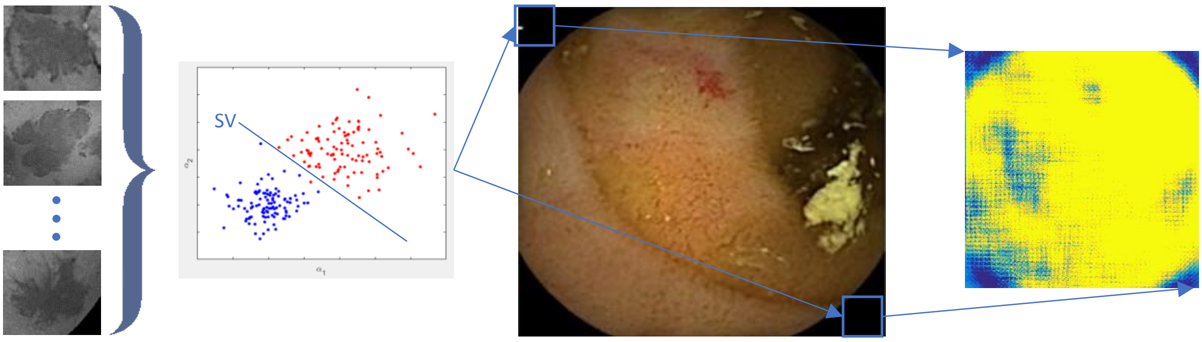

Figure 1: First method consists as a k-means clustering of all the SIFT descriptors extracted from patches containing the type of pathology which are targeted (learning step). In a second step, the learned SVM classifier is computed on the overall test image (sliding window strategy) whih is converted into a probability map using the SVM classification score obtained.

The next step consists in creating the dictionary itself using the all set of SIFT descriptors obtained from the images. In that purpose, we use the classic k-means clustering over the descriptors to perform the classification and obtain the most representative features of the images, which will be the words of the dictionary.

In this section, we want first to evaluate if dictionary learning and bag of features are of related interest for image segmentation. To do so, an interesting idea is to have a classifier that can tell if a histogram of features transposed into the bag of features basis may properly describe an image which belongs to the class of object we want to segment. To achieve so, a two-classes Support Vector Machine (SVM) classifier is a good option because it will provide the class an image belongs to, the targeted object or not, and can gives a score of this classification based on the loss function. This score will be used to have the probability of the image to be part of the targeted object. The all processing scheme leading to the energy map that will be used for active contour segmentation is shown in Fig.1.

At last we associate the heat map with the gray scale image to generate a weighted gray scale image and do the active contour segmentation on the new image coming from this association using the maximization of alpha-divergence, detailed in (Meziou *et al.*, 2012).

2.2 Experiments

In this paper, we worked on small-bowel videocapsule images with a focus on two types of lesions: vascular lesions (angiodysplasia) and inflammatory lesions (including ulcers). The image set we have contains 600 images of angiodysplasia and 450 images of ulcer. It is a part of the CAD-CAP database (Computer-Assisted Detection in Capsule) presented in (Leenhardt et al., 2019). All the images have an associated mask, which is the ground truth manually segmented by physicians and that is used to generate a set of patches containing the pathology (Fig. 2). An equivalent set of negative patches is randomly created by taking square area which does not overlap the ground truth. For all the methods presented in this paper, we used 80% of the patches set as a learning base and kept 20% for the validation process, that is to say to evaluate if the dictionary and SVM classifier leads to a good classification of the testing database. The active contour segmentation process using the learned dictionary is evaluated on two specific images that were not used in the learning process nor in the testing.

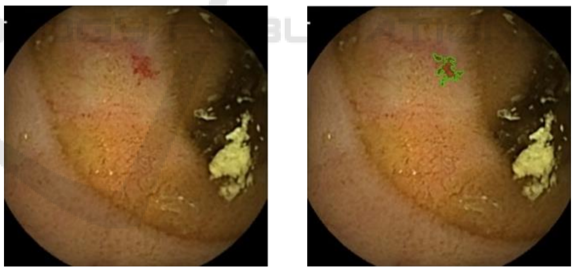

Figure 2: Left: Example of an angiodysplasia image used for testing. Right: The image with the associated mask.

In Fig. 3 and Fig. 4 are shown segmentation results obtained with the classic histogram-based approach of (Meziou et al., 2014). For this segmentation, we choose an initialization of the active contour as a circle inside the targeted object. The segmentation process is designed to work with a detector that can certify the presence of the targeted object and gives an approximate location of it, like the detector proposed in (Angermann et al., 2016). So, the location returned will be used as the center of the initialization circle for the segmentation process.

Figure 3: Segmentation done by the alpha-divergence method based on luminance alone. Red is the result ad green the ground truth segmentation.

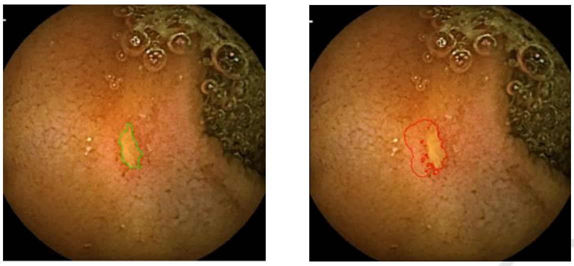

Figure 4: Left: The ulcer image used with the associated segmentation (ground truth). Right: The segmentation done by the alpha-divergence method based on luminance alone.

The segmentation done using the alpha-divergence based on luminance alone, that is to say, the classic approach, doesn't lead to a satisfying result (80% of misclassified pixels) on this image, and does not converge (the process was manually stopped).

Condidering now the first approach proposed in this paper, to constitute the bag of features we need first to define the size of the dictionary (number of clusters for the k-means clustering), if it is obvious to have 256 clusters when using the luminance of pixels like for the alpha-divergence approach, it is not that simpleto determine the optimal size of the considered bag of features. To decide which size would lead to the best representation of the images, we tried several ones and evaluated the SVM classifier trained with the different dictionaries. The results are computed in Table 1 under the form of confusion matrices.

From those results one can notice that all dictionary sizes give high scores for the classification task. Nevertheless we chose the 4000-cluster one as it has the best results in terms of performance.

We propose to try our first method with two different bags of features, using two different dictionaries: The first one is based on chunk of gray scale images using only one patch for a given area presenting with a pathology. For the second, we augmented the size of the patches set by adding chunks of images but with a center shifted by 10%, for each direction, so we have five more patches in

Table 1: Result of the evaluation of the SVM classifier with several dictionary sizes.

Number of clusters	500	Predicted	
		Target	Background
Known	Target	0.91	0.09
	Background	0.10	0.90

Number of clusters	1000	Predicted	
		Target	Background
Known	Target	0.93	0.07
	Background	0.09	0.91

Number of clusters	2000	Predicted	
		Target	Background
Known	Target	0.94	0.06
	Background	0.06	0.94

Number of clusters	4000	Predicted	
		Target	Background
Known	Target	0.97	0.03
	Background	0.06	0.94

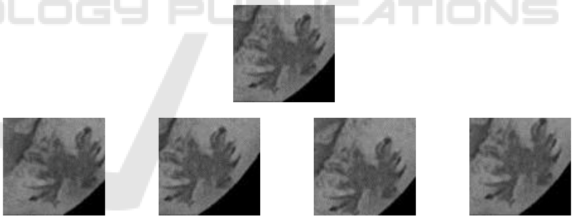

Figure 5: Example of chunk of images, top row is the chunk used for the first method, and bottom row is the four chunks added for the second method (from left to right shifts are left, right, up, down).

the positive set (Fig. 5). We also added the same amount of background images (negative examples) to keep a well balanced image set.

Considering the same initialization scenario as previously introduced for the classic luminance-histogram approach, the first method leads to a satisfying result with less than 30% of misclassified pixels (Fig. 6). But seams less robust that the second method that give stronger results with high probability on the target area and give less probability in the corner of the image where all the pixels are black (Fig. 7). This is from the fact that the second method has more

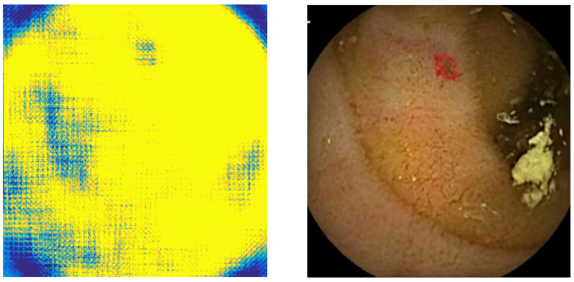

Figure 6: Example of the local probability method, using the first learned dictionary. Left: The probability map. Right: Obtained result at convergence.

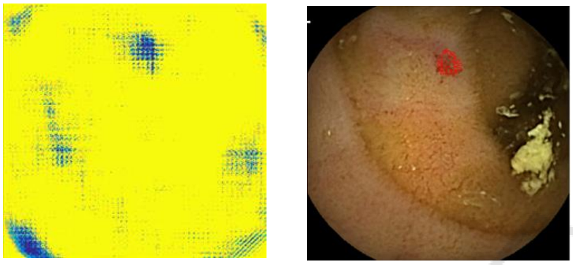

Figure 7: Example of the local probability method, using the second learned dictionary. Left: The probability map. Right: The result.

background features, so it can better recognize them. Additionally, a key point of this method is in the fact that segmentation process has a convergence state, which was not the case with the only luminance based method.

We also challenged our method on a different targeted object, this time we want to segment an ulcer using the second dictionary method with more features extracted from every image from the data set. This example means to prove the robustness of the proposed method. To use the method with another type of targeted object, it is necessary only to change the bag of features dictionary with one providing a good representation of the targeted object. We can see that even if as the probability map seams chaotic, the result of the segmentation is promising (Fig. 8).

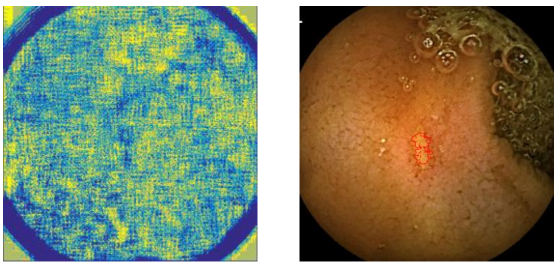

Figure 8: Example of the local probability method on an ulcer, using the second dictionary method. Left: The probability map. Right: The result.

From those results, we know that a method based on bag of features can provide satisfying results. But this method comes with a strong drawback in terms of the computation time: a probability map, in our examples, takes more than 1 hour to be computed for a 500x500 image (on a quad-core 3.3GHz), which is not compatible with a reasonable use of the method.

Moreover, if it demonstrates that Bags of Features can be efficiently used for patch classification, the segmentation process does not use the histogram directly which is not fully satisfying considering the work proposed in (Meziou *et al.*, 2014).

3 ACTIVE CONTOUR OPTIMIZTION USING HISTOGRAMS OF BAG OF FEATURES

3.1 Method

Now that it has been demonstrated that a bag of features can be valuably used in the context of statistical region-based active contour segmentation, we can use them as the base histograms for the alpha-divergence method to improve the statistical representation of the image by replacing the histogram of luminance of the pixels by a histogram of our new bag of features representation of the image.

Here what we want to achieve is to create two references histograms, a histogram that represents the targeted object and another histogram representing what is not the targeted object (background). Those histograms need to be on the same bean-base, so we decided to use the same database we used previously (second dictionary with more features) as we know this bag of features is a good representation for the targeted object and the background.

As introduced previously, the objective is now to use the alpha-divergence approach in a competition scenario, that is to say that the distance between the reference histogram of the targeted object, and the histogram of the inner region of the segmentation and at the same time the distance between the reference histogram for the background and the one of the outer region will be minimized.

This new definition of the statistical representation of the inner and outer regions leads Eq. (6) for the energy J:

$$J(\Gamma, \Omega_{in}, \Omega_{out}) = \int_{\mathfrak{R}} \varphi_\alpha\big(p_{in}, p_{ref\_in}, \lambda_{BoF}\big) d\lambda_{Bo}$$

$$+ \int_{\mathfrak{R}} \varphi_\alpha\big(p_{out}, p_{ref\_out}, \lambda_{BoF}\big) d\lambda_{BoF} + \beta \int_{\Gamma} ds \quad (6)$$

To compute the histograms, we will consider all the images that we manually segmented based on ground truth masks, and we represent them all in the dictionary base. If we take all the object patches and represent them in the dictionary base, we have a reference histogram which represent the occurrences of all the SIFT descriptors for the targeted object. For the background, we do the same but with patches that are not part of the targeted object. Fig. 9 shows the two reference histograms that are going to be used for the alpha-divergence active contour segmentation approach.

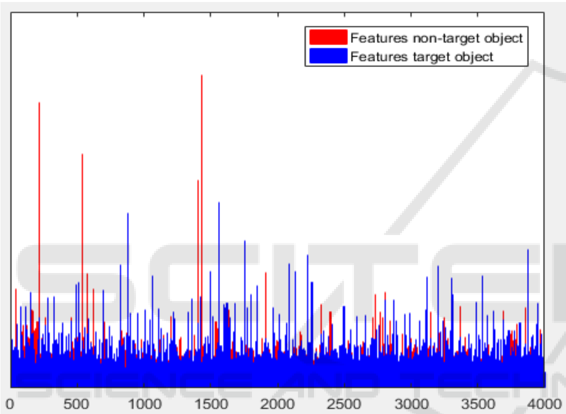

Figure 9: Example of histogram of the features for the target object (blue) and background (red).

Here we can see that the two histograms are different, with for instance the highly-represented features of the background different from the ones for the targeted object.

3.2 Experiments

To implement the use of the histograms of Fig. 9, at each iteration of the evolution process of the active curve, we compute the new histograms for the inner and outer regions and minimize the distance with the references (see Eq. (6)).

Fig. 10 shows segmentation results obtained on the same previously introduced image containing an angiodysplasia. With this new base histograms for the alpha-divergence method, the results are better than with luminance histograms. That method segment 99% of the targeted pixels, and has 45% of misclassified pixels, which is far less than the 80% of the

Figure 10: Example of segmentation done with the alpha-divergence method based on bag of features.

alpha-divergence based on luminance alone, and the results are the same on both types of images. A key point here is the fact that this segmentation converges, unlike the one based on luminance.

The minimization results in a less fit segmentation on the object than the first method because the reference histograms are representations of a “BoF” region that contains the targeted object, so it is not as precise as a pixel-by-pixel computation. Nevertheless, it does not require as much time as the first method to compute.

Fig. 11 shows results obtained on the image containing an inflammatory lesion. Obtained result is quite satisfying and in this case, the final segmentation does not suffer as much as in the previous case of the “mean” effect related to the BoF representation.

Figure 11: Example of segmentation done with the alpha-divergence method based on bag of features.

4 CONCLUSION

In this paper, we proposed a method to improve a statistical-based active contour segmentation using the alpha-divergence family to assess the similarity between two images. The method uses a bag of features and SIFT descriptors as base for the representation of the images. And use references histograms for the targeted object and background of the image, computed from a learned dataset. We try to minimize the distance between the reference histogram of the targeted object and the histogram of the inner region of the segmentation and at the same time the distance between the reference histogram of the background of the image and the histogram of the outer region. This approach provides a good combination of the statistical properties of the whole image. We presented an application of this method on two types of medicals images leading to better results than the luminance base.

As a future work, several approaches can be added to the method, the first one would be to use the optimization of the alpha parameter of the alpha-divergence (Meziou et al., 2014). Another approach can consist in changing the minimization between the histogram of the region and the reference by a maximization of the distance between the histograms of the two regions. It is also possible to investigate further in the statistical representation of the image using more complex representations, as deep features computed from a convolutional neural network.

REFERENCES

Meziou, L., Histace, A., Precioso, P., Matuszewski, B., and Murphy, M., 2011, Confocal Microscopy Segmentation Using Active Contour Based on Alpha-Divergence, *IEEE International Conference on Image Processing (ICIP)*, pp. 3138-3141.

Kass, M., Witkin, A, and Terzopoulos, D., 1988, Snakes: Active contour models, International Journal of Computer Vision, vol. 1, no. 4, pp. 321-331.

Caselles, V., Kimmel, R., and Sapiro, G., 1997, *Geodesic active contours, International Journal of Computer Vision*, vol. 22, no. 1, pp. 61-79.

Chan, T.F., and Vese, L. A., 2001, Active contours without edges, *IEEE Transactions on Image Processing*, vol. 10, pp. 266-277.

Aubert, G., Barlaud, M., Faugeras, O. and Jehan-Besson, S., 2003, Image segmentation using active contours: Calculus of variations or shape gradients?, *SIAM Journal of Applied Mathematics*, vol. 63, pp. 2128-2154.

Meziou, L., Histace, A., Precioso, F., 2014, Statistical region-based active contour using optimization of alpha-divergence family for image segmentation, *IEEE International Conference on Image Processing (ICIP)*, pp. 6066-6070.

Angermann, Q., Histace, A., Romain, O., 2016, Active Learning For Real Time Detection Of Polyps In Videocolonoscopy, *Medical Image Understanding and Analysis Conference*, pp. 182-187.

Meziou, L., Histace, A., Precioso, F., Matuszewski, B. and Carreiras, F., 2012, Fractional Entropy Based Active Contour Segmentation of Cell Nuclei in Actin-Tagged Confocal Microscopy Images, *Medical Image Understanding and Analysis Conference*, pp. 117-123.

Lecellier, F., Jehan-Besson, S., Fadili, J., Aubert, G. and Revenu, M., 2009, Optimization of divergences within the exponential family for image segmentation, *International Conference on Scale Space and Variational Methods in Computer Vision*, Berlin, Heidelberg, 2009, pp. 137–149, Springer-Verlag.

Leenhardt, R., Vasseur, P. Li, C., Rahmi, G., Cholet, F., Saurin, J.-C. …, Histace, A. and Dray, X., 2019, A neural network algorithm for detection of GI angiectasia during small-bowel capsule endoscopy, *Gastrointestinal Endoscopy*,