Contents lists available at ScienceDirect

Pattern Recognition Letters

journal homepage: www.elsevier.com/locate/patrec

Data availability

Links listed below are the data deposition URLs from Data availability section. Please verify the links are valid. This page will not appear in the article PDF file or print. They are displayed in the proof pdf for review purpose only.

https://rvsc.projets.litislab.fr/ Dataset Link: https://rvsc.projets.litislab.fr/
http://www.wisdom.weizmann.ac.il/~vision/Seg_Evaluation_DB/ Dataset Link: http://www.wisdom.weizmann.ac.il/~vision/Seg_Evaluation_DB/
http://prostatemrimagedatabase.com/ Dataset Link: http://prostatemrimagedatabase.com/
https://drive.grand-challenge.org Dataset Link: https://drive.grand-challenge.org

Contents lists available at ScienceDirect

Pattern Recognition Letters

journal homepage: www.elsevier.com/locate/patrec

Highlights

Image segmentation via two-step deep variational priors

Pattern Recognition Letters xxx (xxxx) xxx

Lu Tan*, Xue-Cheng Tai, Ling Li, Wan-Quan Liu, Raymond H. Chan, Dan-Feng Hong

- A two-procedure selective segmentation approach: unsupervised then refinement stages.
- Procedure II implemented as joint deep variational model or as individual unit.
- Mutual benefits achieved by combining variational methods with deep learning.
- Our approach doesn't impose strict requirements on point positions.
- Individual module allows unlimited repetition until corrections reach satisfactory level.

Contents lists available at ScienceDirect

Pattern Recognition Letters

journal homepage: www.elsevier.com/locate/patrec

Image segmentation via two-step deep variational priors

Lu Tan [a,*], Xue-Cheng Tai [b], Ling Li [c], Wan-Quan Liu [d], Raymond H. Chan [c,f], Dan-Feng Hong [g]

[a] School of Computer Science and Information Engineering, Bengbu University, Bengbu, Anhui, China
[b] Norwegian Research Center (NORCE), Bergen, Norway
[c] School of Elec Eng, Comp and Math Sci (EECMS), Curtin University, Perth, Australia
[d] School of Intelligent Systems Engineering, Sun Yat-sen University, Shenzhen, China
[e] Department of Operations and Risk Management and School of Data Science, Lingnan University, Tuen Mun, Hong Kong
[f] Hong Kong Centre for Cerebro-Cardiovascular Health Engineering, Hong Kong
[g] Aerospace Information Research Institute, Chinese Academy of Sciences, Beijing, China

ARTICLE INFO

Editor: Antonio Fernández-Caballero

Dataset link: https://rvsc.projets.litislab.fr/, http://www.wisdom.weizmann.ac.il/~vision/Seg_Evaluation_DB/, http://prostatemrimagedatabase.com/, https://drive.grand-challenge.org

Keywords:
Integration approach
Flexible module
Iterative deep variational priors

ABSTRACT

This paper proposes an iterative deep variational approach for image segmentation in a fusion manner: it is not only able to realize selective segmentation, but can also alleviate the issue of parameter/initialization dependency. Moreover, it possesses a refinement process designed to handle challenging scenarios, such as images containing obscured, damaged, or absent objects, or those with complex backgrounds. Our proposed approach consists of two main procedures, i.e., selective segmentation and shape transformation. The first procedure works as a stem in a totally unsupervised way. A convolutional neural network (CNN) based architecture is properly incorporated into the selective weighting constrained variational segmentation model. The second procedure is to further refine the outputs. This part can be achieved in two ways: one direction is to establish a joint model with the semantic shape constraint. The other technical direction is to make the shape descriptor separated from the joint model and work as an individual unit. In the proposed approach, the minimization problem is transformed from iterative minimization for each variable to automatically minimizing the loss function by learning the generator network parameters. This also leads to a good inductive bias associated with classic variational methods. Extensive experiments have demonstrated the significant advantages.

1. Introduction

Selective segmentation [1,2] is one important technique in image processing. Unlike standard segmentation, it does not require the identification of all objects in an image. As shown in Fig. 1, standard segmentation will segment all four objects in (a), while selective segmentation will only segment the region (c) based on the selected points (by mouse clicks) given in (b).

Although selective image segmentation has received relatively less attention in research, recent advances [3–13] in this field demonstrated its success in medical and real applications. Among the research works, [3–9] adopted variational methods. The research [3–6] centered on exploring effective approaches. [3] proposed a two-step approach for medical images, which included the use of a weighted function and a subsequent thresholding procedure. [4] tried to partition all objects using a global level set function, while segmenting the selected item using a different level set function with a more local focus. Spencer et al. [5] investigated parameter-free selective segmentation to alleviate the user's burden by simplifying the input requirements. [6] employed a newly designed model to smooth the given image and a modified Gout's model [14] to detect the target boundary. [7–9] emphasized on new model design. A penalty term was introduced in [7], which involved the edge-weighted geodesic distance from a marker set. In [8], a model was developed that relied on a region-based approach. It combined edge information and statistical data to effectively capture the desired object, whether the object had a single region or multiple regions. [9] combined the Chan-Vese model with elastica and landmark constraints. As for [10–12], they utilized deep learning techniques. [10] incorporated global contextual information into each local region of interest to enhance feature representation. And a click discounting factor to facilitate the effective end-to-end training of their model. In [11], a CNN architecture was introduced, wherein the four extreme locations were represented as an additional heatmap input channel to the network. The key of [12] lies in the discovery that a coarse-to-fine structure is essential for achieving more accurate

* Corresponding author.
E-mail addresses: lu.tan@bbc.edu.cn (L. Tan), xtai@norceresearch.no (X.-C. Tai), l.li@curtin.edu.au (L. Li), liuwq63@mail.sysu.edu.cn (W.-Q. Liu), raymond.chan@ln.edu.hk (R.H. Chan), hongdf@aircas.ac.cn (D.-F. Hong).

https://doi.org/10.1016/j.patrec.2025.04.030
Received 16 January 2024; Received in revised form 13 April 2025; Accepted 28 April 2025

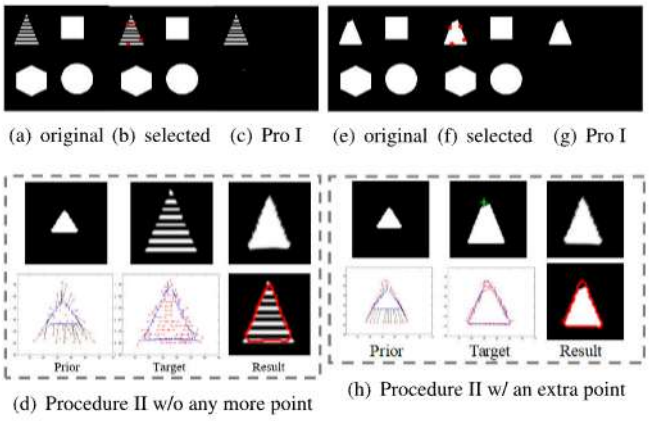


Fig. 1. Our proposed approach for selective shape with a missing part.

segmentation masks, while adding extra layers does not contribute greatly. The thesis work [13] explored the use of variational and deep learning methods for selective segmentation, providing insights into designing models that consider specific criteria in the segmentation output. Current selective segmentation largely focused on segmenting real regions of interest or necessitated precise initialization. In this paper, our proposed approach not only addresses segmentation for objects with actual boundaries but also for objects with illusory or ambiguous boundaries. And some inherent limitations of variational methods are also considered.

As stated in [3,9,15], variational methods are highly effective in producing outputs without sacrificing essential features while also demanding less computational memory. Remarkably, it can be seamlessly integrated with other advanced theories. For example, stochastic programming derived from probability theory [16] has been employed in image processing. However, there exist limitations in the minimization of the energy functional: alternating directional optimization strategy based fast algorithms [9,16–18] is the widely adopted solution, considering both simplification and effectiveness. But they will introduce additional auxiliary variables, leading to an increased number of hyper-parameters that require manual tuning. Besides, these solutions exhibit sensitivity to predefined parameters and dependence on their initial settings.

Among the substantial research conducted by deep learning networks in recent years, convolutional neural networks (CNNs) [19–21] have made significant achievements with promising outcomes. Some work [22] exhibited enhanced segmentation performance while still relying on fully annotated masks as supervision. We believe that the remarkable ability of learning realistic image priors from large data is a crucial factor in CNNs, as learning is the primary driver behind the outstanding performance of deep networks. A good example given in [23] demonstrated that the structure of the network should correspond well with the structure of the data. Similarly, authors in [21] put forward that image statistics can be adequately captured by the structure of a single CNN generator network in an unsupervised way. Good examples [19,24] presented that unified frameworks can be built on top of [21] by coupling multiple CNN generator networks to handle seemingly unrelated computer vision tasks. Taking inspiration from the studies by [25–28], incorporating diverse shape priors into existing deep architectures proves to be a pragmatic and successful approach for improving performance. Additionally, several other studies, including [29–31], have demonstrated the effectiveness of leveraging traditional methods, such as clustering models, in combination with end-to-end representation learning based on specific optimization principles to facilitate segmentation tasks.

Another challenge in segmentation is for images with the occluded/damaged objects or complex backgrounds. To the best of our knowledge, only a limited number of segmentation techniques can address the issue of missing information in such cases. Presently, most existing approaches primarily focus on detecting the visible objects within the image and do not consider whether the missing portions should be reconstructed.

Motivated by these challenges, we propose a novel framework that integrates classical variational functionals with deep networks and deep priors. This integration eliminates the need for traditional optimization algorithms in solving variational functionals while effectively addressing the aforementioned computational difficulties. A key contribution of our work is establishing a meaningful connection between these two approaches, demonstrating their mutual benefits. Specifically, we propose a deep learning method for selective segmentation using a variational framework, with an optional refinement procedure for cases where preliminary results are suboptimal. This refinement capability distinguishes our approach from common selective segmentation methods [3–13]. Our framework uniquely unifies CNNs in a variational way and incorporates prior shape technique without additional annotations. The integration of these three components — variational methods, deep priors, and shape context — is motivated by their complementary strengths: variational methods enable self-supervised modeling of energy functionals with multiple variables, the CNN-based deep image prior (DIP) [21] effectively captures image statistics through a single CNN generator structure, and shape context serves as a highly discriminative descriptor, imposing semantic shape priors during output generation.

Our contributions are briefly summarized as: (1) We propose a novel two-procedure approach by deep variational priors, which can achieve a coarse-to-fine selective segmentation for different situations. (2) It allows to overcome inherent drawbacks existed in variational methods, such as manual hyper-parameter tuning, sensitivity to numerous predefined values, and heavy dependence on initial settings. (3) Our approach does not impose strict requirements on the point positions. Furthermore, it technically permits the unlimited repetition of the individual module until the correction reaches a satisfactory level.

The rest of this paper is structured as: our approach is elaborated in Section 2. Experiments with evaluation, comparisons and analysis are given in Section 3. Section 4 draws the conclusion.

2. The proposed approach

The feasibility and flexibility of our proposed approach lie in: assuming the desired results are obtained from Procedure I, one can directly ignore the refined method in Procedure II. In case that the refined procedure is required, one has two ways to incorporate the shape context based transformation technique: within a joint formulation or an individual module. Procedure II allows it to simultaneously address tasks such as completing missing boundaries, reconstructing occluded object structures, and elevating segmentation accuracy by adding more points. The model and implementation are illustrated in Fig. 2. Our theoretical foundation leverages the connection between variational methods and deep learning. As shown in Fig. 2, let Ω be the image domain, and u or φ: $\Omega \to R$ be the desired segmentation function that minimizes the variational energy. Instead of directly optimizing in the infinite-dimensional function space, we parameterize the solution space using a CNN architecture. Specifically for the segmentation function u, we represent it through a CNN parameterized by θ: $u = DIP_\theta(z)$, where z is the random input. This parameterization is theoretically justified by the Universal Approximation Theorem — given sufficient network capacity, CNNs can approximate any continuous function on a compact domain to arbitrary precision. Formally, for any $\epsilon > 0$, there exists a parameter set θ such that $\|u - DIP_\theta(z)\|_2 < \epsilon$. This guarantees that our CNN-based representation is sufficiently expressive to approximate optimal solutions in the variational framework. For the related work, see Sec. 1 of the Supplementary Material.

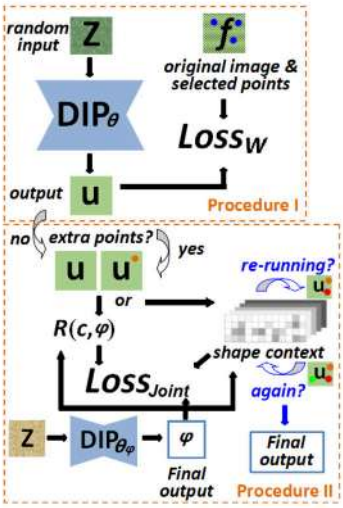


Fig. 2. The modeling and process of our proposed approach.

2.1. Procedure I: Weighted segmentation

As shown in Fig. 2, each variable is obtained by a DIP network instead of being calculated based on its corresponding partial differential equation. With the incorporation of CNN, the original minimization problem of energy functional is transformed into an optimization problem with the new loss below:

$$Loss_W = \alpha|\nabla DIP_\theta(z)| + \omega^2(DIP_\theta(z) - f)^2. \quad (1)$$

From Fig. 2, $u = DIP_\theta(z)$, z is the randomly initialized input (uniform noise) of the DIP and u is the output reconstructed from the original image f with the same size as f. The definition of $\omega^2(x)$ including $d(x)$ and $g(x)$ is given in [3]. ∇ is the gradient operator for smooth output. When $|\nabla(\cdot)|$ is used with DIP, it helps balance detail preservation and noise reduction, improving overall performance. Here, this term is sufficient for Procedure I to obtain desired results, thus $|\nabla(\cdot)|^2$ used in [3,32] is omitted. The architecture details can be found in the Supplementary Material.

2.2. Procedure II: Refined method

Different from the existing work [33,34], we aim to produce refined results for interactive segmentation with the shape matching technique in two modes: (1) if no additional points are provided except the selected ones for the weighting constraint (Section 2.1), shape context is used for enhancing performance. (2) if more points are provided (one or two), shape context can achieve the recovery of occluded/damaged objects. As shown in Fig. 2, if "no", shape context is used for a semantic constraint of shape (if necessary). If "yes", a few more points will be added iteratively to learn the shape prior which can achieve the completion of the object shape. A workflow illustration is available in Section 2 of the Supplementary Material.

Procedure IIA — A joint model with semantic shape constraint: From the full version of our framework in Fig. 2, our proposed refined method can be formulated in a joint model. In particular, the loss function for this deep variational model is:

$$Loss_{Joint} = \underbrace{\gamma|\nabla DIP_{\theta_\varphi}(z)| + R(c, DIP_{\theta_\varphi}(z))}_{Loss_{Seg}} + \underbrace{C(DIP_{\theta_\varphi}(z), \Phi_{Prior})DIP_{\theta_\varphi}(z)}_{Loss_{Shape}}, \quad (2)$$

where $Loss_{seg}$ maintains geometric constraints and $Loss_{Shape}$ enforces semantic priors, making the optimization well-posed. $\varphi = DIP_{\theta_\varphi}(z)$ and $C(target, prior) = \sum_i C(target_i, prior_{\eta(i)})$ ($\eta(i)$ is a permutation) refers to the matching from target to prior, which is achieved via minimization. Target should contain the extra points beyond the ones from Procedure I if available. And $R(c, DIP_{\theta_\varphi}(z))$ in (2) is

$$R(c, DIP_{\theta_\varphi}(z)) = \alpha_1(c_1 - f)^2 DIP_{\theta_\varphi}(z) + \alpha_2(c_2 - f)^2(1 - DIP_{\theta_\varphi}(z)), \quad (3)$$

which keeps the same definition as in [16,19], a binary representation for the mask is required, and the advantage is that we can achieve this automatically with no more need for the extra regularization loss as used in [19] and the threshold method used in [16]. Moreover, it naturally gives the knowledge (c_1, and c_2 are dynamically updated) for image foreground and background. Thus the extra optimization procedure to guarantee fore/back-ground as did in [19] is not required.

Procedure IIB — An individual module: Beyond the joint model proposed above, the shape transformation can work as an individual module. After this separation, it can be deduced that (2) will naturally reduce into a pure segmentation approach: $Loss_{Seg} = \gamma|\nabla DIP_{\theta_\varphi}(z)| + R(c, DIP_{\theta_\varphi}(z))$. This reduced version of (2) can also be replaced by the simple thresholding used in [3] or any other different threshold by trial and error. Based on the binary results by thresholding, the shape transformation is then applied with/without extra points given for refinement.

In summary, a two-step segmentation approach is proposed based on deep neural networks and variational approaches. By utilizing useful regularization terms and semantic shape transformations, we can provide valuable guidance for deep neural network design and establish new constraints. Moreover, the proposed formulations maintain the mathematical properties of the original variational problem while leveraging the CNN's representation power. In implementation, we employed multiple strategies to improve convergence: (1) random initializations of network parameters, (2) the Adam optimizer to better handle saddle points, and (3) repeated training attempts to select the best performing model. While these approaches enhanced the optimization process, gradient descent on θ can achieve better convergence behavior, though convergence to a favorable local minimum remains probabilistic rather than guaranteed.

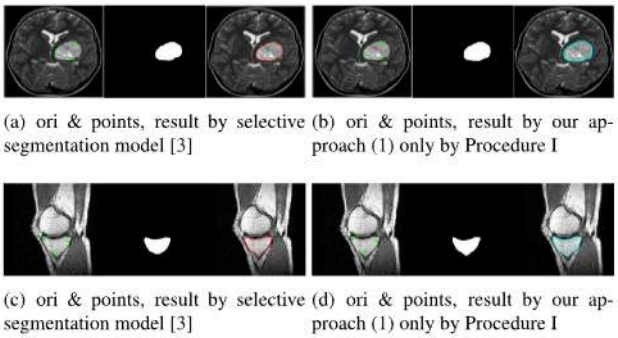
3. Experiments

Experiments are conducted using GTX 1050Ti GPU. Synthetic and natural images (resized into 256 × 256 or 512 × 512) are set as testing images, which are synthesized or chosen from public datasets of MPEG-7, BSD500, PASCAL, RVSC and Weizmann[1]. In Section 3.4, the training dataset is from PROMISE12 challenge and the testing dataset is from the Brigham and Women's Hospital[2]. In our implementation, we set the weighting parameters α and γ in Eqs. (1) and (2) to 0.1, while α_1 and α_2 in Eq. (3) are set to 1 by default. The weighting parameters α and γ were empirically determined. While these parameters are not critical to the core contribution of our method, we set them through experimental exploration. With these settings, our model demonstrated stable behavior and effective feature extraction capability. Similar performance can be achieved with a reasonable range of parameter values, indicating the robustness of our approach.

3.1. Selective segmentation only using Procedure I

Several experimental examples obtained by selective segmentation model [3], double-DIP [19] and our proposed model (1) are given. In Fig. 3, the original images, selected points used by [3] and results obtained by [3] are presented in (a) and (c). (b) and (d) show the

[1] http://www.wisdom.weizmann.ac.il/~vision/Seg_Evaluation_DB/
[2] Prostate MR image database, http://prostatemrimagedatabase.com/



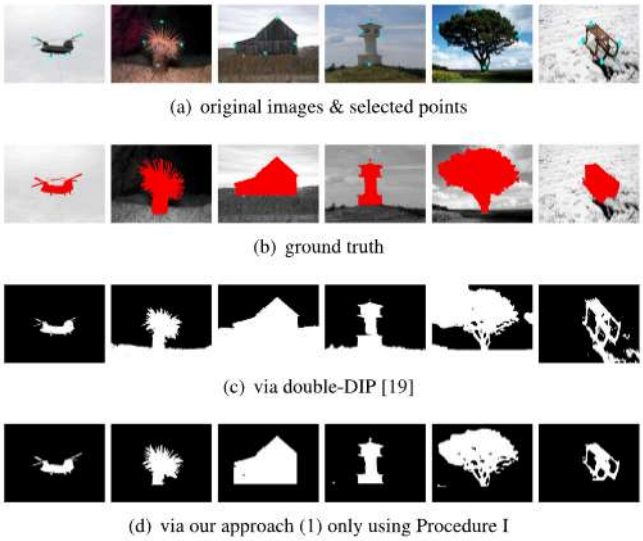
(a) ori & points, result by selective segmentation model [3]

(b) ori & points, result by our approach (1) only by Procedure I

(c) ori & points, result by selective segmentation model [3]

(d) ori & points, result by our approach (1) only by Procedure I

Fig. 3. Comparison between model [3] and our approach (1)-Procedure I. Please note that images and results in (b) and (f) were presented in [3].



(a) original images & selected points

(b) ground truth

(c) via double-DIP [19]

(d) via our approach (1) only using Procedure I

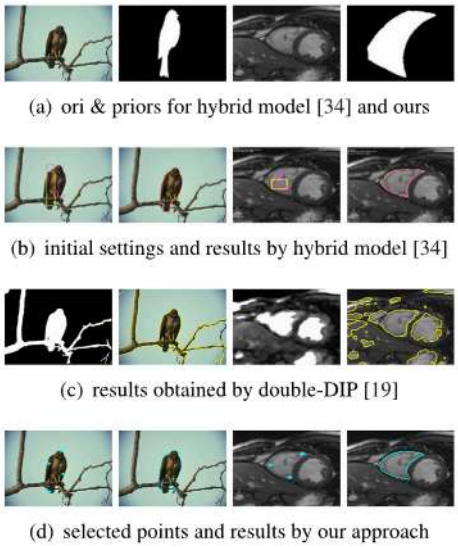
Fig. 4. Comparison between double-DIP [19] and our proposed approach (1). Here images of (c) were directly taken from [19].

related ones of our approach (1) only using Procedure I. We choose the same point locations for a fair comparison. Compared between (a) and (b), competitive performance can be achieved by our model (1) by only using Procedure I. From (c) and (d), it can be observed that our model can capture more detailed information, especially on preserving the corner. However, the three corners in (c) are smeared. Furthermore, our approach requires significantly less computational overhead and minimal hyper-parameter tuning compared to the existing method [3]. While these improvements might appear subtle at first glance, they represent significant advances in detail preservation, achieved with markedly lower computational complexity and fewer hyper-parameters to tune.

In Fig. 4, the complicated background causes difficulty for both [19] and our approach. In our approach, an approximate region of interest can be provided in advance through the interactive segmentation technique, then the background that may affect the final result is excluded. This simple technique can also benefit [19], but [19] is not a good backbone for interactive segmentation since prior information for foreground and background is needed in advance and should be added to the loss of the first optimization (two optimizations in total) for stable segmentation. That would lead to extra effort. On the contrary, this preliminary procedure is not necessary for our model (1).

Then our comparison with double-DIP [19] has been strengthened through a comprehensive evaluation framework. In Table 1, we present detailed quantitative metrics across diverse image types, including:

- Natural objects (chopper: FM 0.954 vs 0.865)



(a) ori & priors for hybrid model [34] and ours

(b) initial settings and results by hybrid model [34]

(c) results obtained by double-DIP [19]

(d) selected points and results by our approach

Fig. 5. Comparison with the hybrid model [34], double-DIP [19] and ours. Images, priors and results shown in (b) were taken from [34].

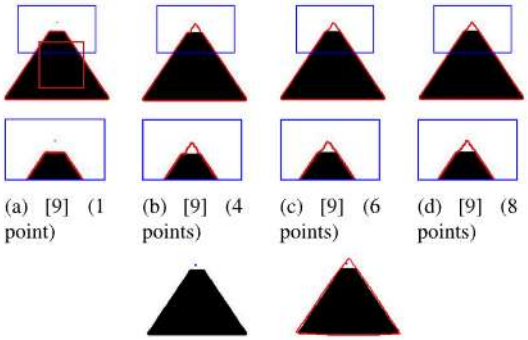
- Textured scenes (tendrils: FM 0.894 vs 0.457)
- Architectural images (house: FM 0.901 vs 0.553, tower: FM 0.957 vs 0.511)
- Organic shapes (tree: FM 0.902 vs 0.734)
- Complex backgrounds (snow: FM 0.734 vs 0.660)

This systematic evaluation demonstrates our method's consistent superior performance across different image categories, while also requiring significantly fewer computational steps (average 458 vs 5583) and less processing time (average 59.9s vs 429.6s). The improved FM scores and computational efficiency validate our approach's robustness across diverse visual scenarios.

Observed from the experimental results on medical and natural images presented in this part, our proposed Procedure I model (1) gives promising performance for selective segmentation. Although our proposed model (1) with only Procedure I cannot capture any semantics, it is still able to obtain high-quality interactive segmentation without the need of the foreground and background information in a totally unsupervised way. There will be no need for any refinement when desired outputs have been produced by our proposed model (1). For semantic preserving capability, satisfactory outputs via Procedure II will be given in the following subsections.

3.2. Interactive segmentation without extra points

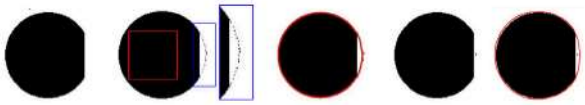
Here the performance of our interactive segmentation approach with semantic preservation is presented. The results obtained by hybrid model with semantic constraint [34] using pure variational methods and unsupervised double-DIP model [19] with pure deep learning approach are used for comparison. In Fig. 5, original images, shape priors used by hybrid model [34] and our proposed approach are given in (a). (b), (c) and (d) display the results obtained by the hybrid model [34], double-DIP [19] and our approach. The results show that our approach could obtain competitive performance compared with the hybrid model [34]. In addition, the high computational cost in one iteration is still kept as described in [34]. As it still needs to design variational based algorithms for efficiency improvement, the inherent problems of parameter increase and parameter sensitivity are inevitable in the hybrid model [34]. But these drawbacks will not impact our work since the introduction of CNN architecture completely changes the traditional optimization. Furthermore, random input is used by our deep variational framework, the necessary contour initialization in the



(a) [9] (1 point) (b) [9] (4 points) (c) [9] (6 points) (d) [9] (8 points)

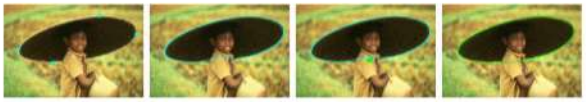
(e) our approach (only 1 point)

Fig. 6. Comparison between CVEL [9] and our approach for object with a small damage. Images and results in (a)–(d) were presented in [9].



(a) original (b) points used for CVEL [9] (c) CVEL [9] (d) our approach (only 1 point)

Fig. 7. Comparison between CVEL [9] and our approach for object with a large damage. Images and results in (a)–(c) were taken from [9].



(a) 1 correction

(b) repeatable correction in detail

Fig. 8. The whole process of our repeatable refinement.

hybrid model [34] (which is also a traditional process in variational methods) shown in Fig. 5(b) no longer exists. Apart from these inherited defects, [34] can only accomplish the task discussed in this subsection. There is no other correction or refinement designed for occlusions or damages.

As for the related work double-DIP [19], since it can only achieve segmentation task without the use of semantic cues, it is not able to extract the region of interest as our proposed approach achieves. As far as we know, our proposed approach is the first attempt for deep unsupervised learning to achieve interactive segmentation augmented with semantic shape prior, making it applicable in a wide range of scenarios. Quantitative comparisons are shown in Tables 1–4.

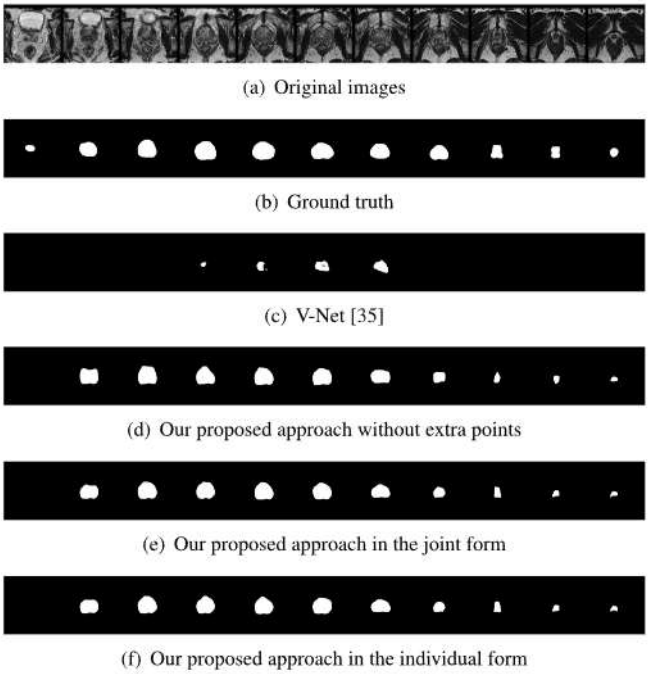
3.3. Interactive segmentation with extra points

3.3.1. Experiments by Procedure I & II-joint form

Objects with damages: The Chan-Vese model with Elastica and Landmark constraints (CVEL) [9] is used for comparison with our model in Fig. 6. (a)–(d) show the results obtained by different landmark points via CVEL model [9]. (e) presents the results by approach. Red rectangle gives the initialized contour, and blue ones present zoomed regions for observing clear differences. There are some limitations in CVEL model [9] based on Fig. 6: landmark points are required to be exact on the object boundaries and corners, such strict conditions

Table 1
Varieties between deep neural networks [19], variational methods [34] and our proposed approach on accuracy and efficiency.

double-DIP [19] — Fig. 4 (c)				
Image	FM	Steps	Time (s)	Time/Steps
chopper	0.865	6000	467.3	0.078
tendrils	0.457	5500	401.9	0.073
house	0.553	6000	474.2	0.079
tower	0.511	5500	423.5	0.077
tree	0.734	6000	468.3	0.078
snow	0.660	4500	342.2	0.076
Average	0.630	5583	429.6	0.077
Ours (Procedure I) — Fig. 4 (d)				
Image	FM	Steps	Time (s)	Time/Steps
chopper	0.954	300	39.7	0.132
tendrils	0.894	400	52.3	0.130
house	0.901	300	31.8	0.106
tower	0.957	350	46.2	0.132
tree	0.902	1000	136.1	0.136
snow	0.734	400	53.2	0.133
Average	0.890	458	59.9	0.128
hybrid model [34] — Fig. 5 (b)				
Image	FM	Steps	Time (s)	Time/Steps
bird	0.973	20	101.7	5.09
medical	0.938	15	34.5	2.30
Average	0.956	17.5	68.1	3.70
double-DIP [19] — Fig. 5 (c)				
Image	FM	Steps	Time (s)	Time/Steps
bird	0.682	6000	448.1	0.075
medical	0.702	6000	456.2	0.076
Average	0.692	6000	452.2	0.076
Ours — Fig. 5 (d)				
Image	FM	Steps	Time (s)	Time/Steps
bird	0.915	700	70.7	0.10
medical	0.952	600	65.3	0.11
Average	0.934	650	67.8	0.105



(a) Original images

(b) Ground truth

(c) V-Net [35]

(d) Our proposed approach without extra points

(e) Our proposed approach in the joint form

(f) Our proposed approach in the individual form

Fig. 9. Different segmentation approaches on Patient 46.

usually consume much more time for convergence. Good performance of the CVEL model [9] is decided by giving more points. However, [9]

Table 2
Varieties between variational methods and ours.

Varieties	Sel. seg [3]	Hybrid [34]	CVEL [9]	Ours
No init.	✓	×	×	✓
No reinit.	✓	×	✓	✓
Parameter	no fixed	no fixed	no fixed	fixed
Semantic	×	✓	×	✓
Refine	×	×	✓	✓
Point num	N/A	N/A	8 (Fig. 6) 17 (Fig. 7)	1 (Fig. 6) 1 (Fig. 7)

Table 3
Varieties between deep architectures and ours.

Varieties	dbl.-DIP [19]	I-O model [12]	Ours
No hint	×	N/A	✓
CNN num	3	N/A	1(seg),2(refine)
Unsupervised	✓	×	✓
Refine	×	✓	✓
Shape compl.	N/A	N/A	✓

Table 4
Evaluation metrics on (a) V-Net [35], (b) our approach without extra points, (c) our approach-joint form, (d) our approach-individual form.

Metrics	V-Net	Ours (no extra)	Ours (joint)	Ours (indiv.)
Dice Coeff.	0.2452	0.7870	0.7968	0.7940
Jaccard Coeff.	0.1397	0.6488	0.6622	0.6584
VS	0.1400	0.7108	0.7005	0.6971
Adj. Rand Idx	0.1664	0.6331	0.6475	0.6496
AUC	0.4987	0.5793	0.5714	0.5690
Mahala. Dist.	1.1694	0.6456	0.6862	0.6837
Haussdorff Dist.	3.2161	1.3632	1.3379	1.2274

fails to reconstruct the contour with only one point provided. On the contrary, much fewer points with approximate positions are sufficient for our proposed approach to producing desired outputs. (e) shows an extreme situation: one point with an estimated location, the final contour is reconstructed.

In Fig. 7, one example of an object with large damage is given. The damaged circle, initialized contour with key landmarks and results obtained by the CVEL model [9] are shown in (a)–(c). (d) gives the results obtained by our proposed approach. Through the comparison of extra points provided as well as the results in (c) and (d), we see that much more points with exact locations on the edge are required for the CVEL model [9] to tackle large damages. For this case, we still use only one point.

3.3.2. Experiments by Procedure I & II-individual form

In Procedure II, the difference between the two types of fusing semantic shape prior is the dynamical adjustment in each refinement process. Individual unit allows to repeatedly correct the final result rather than redo the whole thing from the beginning. In Fig. 8, two examples in real life are used to introduce the whole process of repeatable refinement of our proposed approach. It is clear that the repeatable power of our approach can produce more accurate outputs with the help of the iterative interactive segmentation, and shape transformation.

As for work proposed in [12], maybe more points are used by our approach, but the inside-outside model [12] relies on fully labeled samples, which means it cannot work well based on very few training samples or when only a few simple shape priors are provided. Although there is a stage in [12] for further correction, its excellent performance primarily centers on capturing actual object boundaries in the image. It cannot achieve shape completion of objects with occluded/damaged/missing parts as the Procedure II of our proposed work can do.

3.4. Experiments on objects with vague areas/obscure edges

Experiments on prostate MRI with limited resolution and low-quality problems are conducted in this part. A comparison with the V-Net [35] on Patient 46 of the test dataset from the Brigham and Women's Hospital is provided. The training dataset of PROMISE12 with 50 patients is used to train the V-Net. Among metrics for evaluating the performance of segmentation approaches, we selected six prevalent ones: overlap-based *Dice Coefficient*, *Jaccard Coefficient*, volume-based *Volumetric Similarity (VS)*, pair-counting-based *Adjusted Rand Index*, probabilistic-based *Area Under the ROC Curve (AUC)*, *Mahalanobis Distance* and *Haussdorff Distance* (HD_{95}).

In Fig. 9, the original MRI images as well as the corresponding ground truth are presented in the first two rows. The last four rows show results from V-Net [35], our proposed approach without extra points, our proposed approach in the joint form and our proposed approach in the individual form. Observations from these selected examples of 2D visualization demonstrate that our segmentation can achieve better prostate shape reconstruction from the whole image. The quantitative comparison between our proposed fusion approach and the classical V-Net model is presented in Table 4, using six evaluation metrics. We chose V-Net as a representative supervised learning framework for this comparison, as it demonstrates how our unsupervised method can effectively complement supervised approaches. While more recent supervised architectures exist, the complementary principles demonstrated with V-Net would apply similarly to these newer models. The experimental results show that our method achieves high accuracy, making it particularly valuable in scenarios with limited training data or when additional refinement is needed beyond the initial supervised segmentation.

Regarding joint form vs individual form choice: the empirical comparison between joint and individual formulations reveals comparable performance quality, as demonstrated in Fig. 9. This equivalence in outcomes offers practitioners implementation flexibility tailored to their specific requirements. The joint formulation facilitates end-to-end optimization, while the individual approach enables modular refinement of components. The selection between these formulations can be guided by application-specific constraints, particularly when considering the necessity of iterative refinement in the development pipeline. For instance, as illustrated in Fig. 8(b), when multiple iterative refinements are needed, the individual form offers clear advantages — it allows for localized adjustments and step-by-step refinements without recomputing the entire model, which would be computationally intensive in the joint form. This flexibility in the individual form is particularly valuable for applications requiring fine-tuning of specific regions or multiple refinement iterations. In contrast, the joint form may be more suitable for applications requiring one-time, global optimization. More discussion are in Sec. 3 of the Supplementary Material.

4. Conclusion

This paper presents a two-step deep variational framework that successfully combines CNN architectures with variational methods for image segmentation. Our approach offers several key strengths: the CNN-based dynamic generator enhances image representation while reducing computational complexity and parameter sensitivity; the two-step mechanism effectively handles challenging scenarios including missing boundaries and occluded structures; and the variational energy functional provides proper guidance for architecture parameter tuning. However, our approach does face certain limitations in fine detail segmentation scenarios, particularly with retinal vessel images. In these cases, Procedure I tends to misidentify detailed structures as noise, resulting in disconnected or missing vessel segments, while Procedure II struggles with generating appropriate priors for non-uniform, intricate structures, making shape prior generation computationally expensive and often ineffective. Despite these challenges, extensive experimental results demonstrate that our method achieves comparable or superior performance to state-of-the-art approaches, especially in cases with incomplete or obscured object boundaries.

CRediT authorship contribution statement

Lu Tan: Writing – review & editing, Writing – original draft, Visualization, Software, Project administration, Methodology, Formal analysis, Conceptualization. **Xue-Cheng Tai:** Writing – review & editing, Supervision, Funding acquisition. **Ling Li:** Writing – review & editing, Formal analysis, Data curation. **Wan-Quan Liu:** Writing – review & editing, Methodology, Data curation, Conceptualization. **Raymond H. Chan:** Writing – review & editing, Formal analysis. **Dan-Feng Hong:** Writing – review & editing.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

This work was supported in part by the NORCE Kompetanseoppbygging Program; in part by the Guangdong Province Pearl River Leading Talents Program (2021CX02G450); in part by HKRGC Grants (C1013-21GF, LU11309922) and ITF Grants (MHP/054/22, LU BGR 105824).

Appendix A. Supplementary data

Supplementary material related to this article can be found online at https://doi.org/10.1016/j.patrec.2025.04.030.

Data availability

The datasets used in this study are publicly available from MPEG-7, BSD500, PASCAL, RVSC (https://rvsc.projets.litislab.fr/), Weizmann (http://www.wisdom.weizmann.ac.il/~vision/Seg_Evaluation_DB/), PROMISE12, Brigham and Women's Hospital (http://prostatemrimagedatabase.com/), and DRIVE (https://drive.grand-challenge.org).

References

[1] Noor Badshah, Ke Chen, Image selective segmentation under geometrical constraints using an active contour approach, Commun. Comput. Phys. 7 (4) (2010) 759.
[2] Jack Spencer, Ke Chen, A convex and selective variational model for image segmentation, Commun. Math. Sci. 13 (6) (2015) 1453–1472.
[3] Chunxiao Liu, Michael Kwok-Po Ng, Tieyong Zeng, Weighted variational model for selective image segmentation with application to medical images, Pattern Recognit. 76 (2018) 367–379.
[4] Afzal Rahman, Haider Ali, Noor Badshah, Lavdie Rada, Ayaz Ali Khan, Hameed Hussain, Muhammad Zakarya, Aftab Ahmed, Izaz Ur Rahman, Mushtaq Raza, Muhammad Haleem, A selective segmentation model using dual-level set functions and local spatial distance, IEEE Access 10 (2022) 22344–22358.
[5] Jack Spencer, Ke Chen, Jinming Duan, Parameter-free selective segmentation with convex variational methods, IEEE Trans. Image Process. 28 (5) (2019) 2163–2172.
[6] Wenxiu Zhao, Weiwei Wang, Xiangchu Feng, Yu Han, A new variational method for selective segmentation of medical images, Signal Process. 190 (2022) 108292.
[7] Michael Roberts, Ke Chen, Klaus Loureiro Irion, A convex geodesic selective model for image segmentation, J. Math. Imaging Vision 61 (2018) 482–503.
[8] Haider Ali, Shah Faisal, Ke Chen, Lavdie Rada, Image-selective segmentation model for multi-regions within the object of interest with application to medical disease, Vis. Comput. 37 (2021) 939–955.
[9] Jintao Song, Huizhu Pan, Wanquan Liu, Zisen Xu, Zhenkuan Pan, The chanvese model with elastica and landmark constraints for image segmentation, IEEE Access 9 (2020) 3508–3516.
[10] JunHao Liew, Yunchao Wei, Wei Xiong, Sim-Heng Ong, Jiashi Feng, Regional interactive image segmentation networks, in: 2017 IEEE International Conference on Computer Vision, ICCV, IEEE Computer Society, 2017, pp. 2746–2754.
[11] Kevis-Kokitsi Maninis, Sergi Caelles, Jordi Pont-Tuset, Luc Van Gool, Deep extreme cut: From extreme points to object segmentation, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2018, pp. 616–625.
[12] Shiyin Zhang, Jun Hao Liew, Yunchao Wei, Shikui Wei, Yao Zhao, Interactive object segmentation with inside-outside guidance, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2020, pp. 12234–12244.
[13] Liam Burrows, Variational and Deep Learning Methods for Selective Segmentation (Ph.D. thesis), The University of Liverpool (United Kingdom), 2022.
[14] Christian Gout, Carole Le Guyader, Luminita Vese, Segmentation under geometrical conditions using geodesic active contours and interpolation using level set methods, Numer. Algorithms 39 (2005) 155–173.
[15] Lu Tan, Zhenkuan Pan, Wanquan Liu, Jinming Duan, Weibo Wei, Guodong Wang, Image segmentation with depth information via simplified variational level set formulation, J. Math. Imaging Vision 60 (1) (2018) 1–17.
[16] Lu Tan, Ling Li, Wanquan Liu, Jie Sun, Min Zhang, A novel Euler's elastica-based segmentation approach for noisy images using the progressive hedging algorithm, J. Math. Imaging Vision 62 (1) (2020) 98–119.
[17] Liang-Jian Deng, Roland Glowinski, Xue-Cheng Tai, A new operator splitting method for the Euler elastica model for image smoothing, SIAM J. Imaging Sci. 12 (2) (2019) 1190–1230.
[18] Lu Tan, Wanquan Liu, Zhenkuan Pan, Color image restoration and inpainting via multi-channel total curvature, Appl. Math. Model. 61 (2018) 280–299.
[19] Yosef Gandelsman, Assaf Shocher, Michal Irani, Double-DIP: Unsupervised image decomposition via coupled deep-image-priors, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2019, pp. 11026–11035.
[20] Chunwei Tian, Yong Xu, Wangmeng Zuo, Image denoising using deep CNN with batch renormalization, Neural Netw. 121 (2020) 461–473.
[21] Dmitry Ulyanov, Andrea Vedaldi, Victor Lempitsky, Deep image prior, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2018, pp. 9446–9454.
[22] Shilpa Gite, Abhinav Mishra, Ketan Kotecha, Enhanced lung image segmentation using deep learning, Neural Comput. Appl. (2022) 1–15.
[23] Chiyuan Zhang, Samy Bengio, Moritz Hardt, Benjamin Recht, Oriol Vinyals, Understanding deep learning requires rethinking generalization, in: Proceedings of the International Conference on Learning Representations, 2017.
[24] Lu Tan, Ling Li, Wan-Quan Liu, Sen-Jian An, Kylie Munyard, Unsupervised learning of multi-task deep variational model, J. Vis. Commun. Image Represent. 87 (2022) 103588.
[25] Fei Chen, Huimin Yu, Roland Hu, Xunxun Zeng, Deep learning shape priors for object segmentation, in: 2013 IEEE Conference on Computer Vision and Pattern Recognition, 2013, pp. 1870–1877.
[26] Jun Liu, Xiangyue Wang, Xue-Cheng Tai, Deep convolutional neural networks with spatial regularization, volume and star-shape priors for image segmentation, J. Math. Imaging Vision 64 (6) (2022) 625–645.
[27] Varun Vasudevan, Maxime Bassenne, Md Tauhidul Islam, Lei Xing, Image classification using graph neural network and multiscale wavelet superpixels, Pattern Recognit. Lett. 166 (2023) 89–96.
[28] Yongzhe Yan, Stefan Duffner, Xavier Naturel, Anthony Berthelier, Christophe Garcia, Christophe Blanc, Thierry Chateau, Two-stage human hair segmentation in the wild using deep shape prior, Pattern Recognit. Lett. 136 (2020) 293–300.
[29] James Liang, Tianfei Zhou, Dongfang Liu, Wenguan Wang, CLUSTSEG: Clustering for universal segmentation, 2023, ArXiv.
[30] Chen Liang, Wenguan Wang, Jiaxu Miao, Yi Yang, GMMSeg: Gaussian mixture based generative semantic segmentation models, in: Advances in Neural Information Processing Systems, 2022.
[31] Liulei Li, Tianfei Zhou, Wenguan Wang, Jianwu Li, Yi Yang, Deep hierarchical semantic segmentation, in: 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition, CVPR, 2022, pp. 1236–1247.
[32] Xiaohao Cai, Raymond Chan, Tieyong Zeng, A two-stage image segmentation method using a convex variant of the Mumford–Shah model and thresholding, SIAM J. Imaging Sci. 6 (1) (2013) 368–390.
[33] Serge Belongie, Jitendra Malik, Jan Puzicha, Shape matching and object recognition using shape contexts, IEEE Trans. Pattern Anal. Mach. Intell. 24 (4) (2002) 509–522.
[34] Bin Wang, Xiuying Yuan, Xinbo Gao, Xuelong Li, Dacheng Tao, A hybrid level set with semantic shape constraint for object segmentation, IEEE Trans. Cybern. 49 (5) (2018) 1558–1569.
[35] Fausto Milletari, Nassir Navab, Seyed-Ahmad Ahmadi, V-net: Fully convolutional neural networks for volumetric medical image segmentation, in: 2016 Fourth International Conference on 3D Vision, 3DV, IEEE, 2016, pp. 565–571.