CSIAM Trans. Appl. Math.
doi: 10.4208/csiam-am.SO-2021-0049

Vol. **3**, No. 4, pp. 662-691
December 2022

Unsupervised Deep Learning Meets Chan-Vese Model

Dihan Zheng[1], Chenglong Bao[1,2,*], Zuoqiang Shi[3,2], Haibin Ling[4] and Kaisheng Ma[5]

[1] *Yau Mathematical Sciences Center, Tsinghua University, China.*
[2] *Yanqi Lake Beijing Institute of Mathematical Sciences and Applications, China.*
[3] *Department of Mathematical Sciences, Tsinghua University, China.*
[4] *Department of Computer Sciences, Stony Brook University, USA.*
[5] *Institute for Interdisciplinary Information Sciences, Tsinghua University, China.*

Received 26 November 2021; Accepted 10 April 2022

Abstract. The Chan-Vese (CV) model is a classic region-based method in image segmentation. However, its piecewise constant assumption does not always hold for practical applications. Many improvements have been proposed but the issue is still far from well solved. In this work, we propose an unsupervised image segmentation approach that integrates the CV model with deep neural networks, which significantly improves the original CV model's segmentation accuracy. Our basic idea is to apply a deep neural network that maps the image into a latent space to alleviate the violation of the piecewise constant assumption in image space. We formulate this idea under the classic Bayesian framework by approximating the likelihood with an evidence lower bound (ELBO) term while keeping the prior term in the CV model. Thus, our model only needs the input image itself and does not require pre-training from external datasets. Moreover, we extend the idea to multi-phase case and dataset based unsupervised image segmentation. Extensive experiments validate the effectiveness of our model and show that the proposed method is noticeably better than other unsupervised segmentation approaches.

AMS subject classifications: 68U10, 68T45

Key words: Image segmentation, Chan-Vese model, variational inference, unsupervised learning.

---

1 Introduction

Image segmentation is one of the fundamental problems in image processing and has many applications in computer vision such as object detection [16], recognition [43] and

*Corresponding author. *Email addresses:* `clbao@mail.tsinghua.edu.cn` (C. Bao), `zhengdh@mails.tsinghua.edu.cn` (D. Zheng), `zqshi@tsinghua.edu.cn` (Z. Shi), `haibin.ling@stonybrook.edu` (H. Ling), `kaisheng@tsinghua.edu.cn` (K. Ma)

medical image analysis [35]. Despite great improvements in image segmentation in recent years, it remains challenging and deserves further exploration. Specifically, given an open and bounded domain $\Omega \subset \mathbb{R}^2$ and an image $I:\Omega \to \mathbb{R}$, segmentation aims at finding a decomposition of the region $\Omega = (\cup_{i=1}^N \Omega_i) \cup \Gamma$, where Γ is the closed segmentation curve and $\Omega_i, i=1,2,\cdots,N$ are disjoint open regions of interests. One seminar work is the so-called Mumford-Shah model [29] that minimizes the following functional:

$$E_{\rm MS}(J,\Gamma) := \int_\Omega (I-J)^2 dx + \mu \int_{\Omega \setminus \Gamma} |\nabla J|^2 dx + \nu |\Gamma|, \tag{1.1}$$

where J is a piecewise smooth approximation of I and $|\Gamma|$ is the length of Γ. Here, μ and ν are two positive constants and $|\Gamma|$ can be written as $\mathcal{H}^1(\Gamma)$, which is the 1-dimensional Hausdorff measure. The difficulty in studying (1.1) is that it involves two unknowns J and Γ of different natures: J is a function defined on a 2-dimensional space, while Γ is a 1-dimensional set. It is not easy to minimize the Mumford-Shah functional $E_{\rm MS}$ and the simplified Chan-Vese (CV) model [11] is proposed by using the piecewise constant assumption and thus the functional is reduced to:

$$E_{\rm CV}(c_1,c_2,\Omega) = \int_{\Omega_1} (I-c_1)^2 dx + \int_{\Omega \setminus \Omega_1} (I-c_2)^2 dx + \nu |\partial \Omega_1|, \tag{1.2}$$

where c_1,c_2 are constants for foreground (fg) and background (bg) respectively and $\partial\Omega_1$ is the boundary of Ω_1. Throughout this paper, we called Ω_1 as the fg and $\Omega \setminus \Omega_1$ as the bg. Also, under the maximum a posterior (MAP) framework, the Chan-Vese model (1.2) is derived in [14] by assuming that fg/bg are random variables that generated from two Gaussian distributions and the prior distribution of the boundary is the length regularization term in (1.2). The Gaussian distribution assumptions for fg/bg are key factors for the success of segmentation but they do not hold for complex scenes. In Fig. 1, we illustrate two typical cases: (i) The fg or bg does not satisfy the Gaussian distribution hypothesis (see the first row in Fig. 1(b)); (ii) the distributions of fg and bg have been significantly overlapped (see the second row in Fig. 1(b)). The results of the CV model are present in Fig. 1(c). It is clear that the CV model fails in both cases due to the violation of its basic assumption. To widen the CV model's application range, there is a need to construct a more accurate model applicable to complex scenes.

In recent years, deep learning methods have achieved state-of-the-art performance in image segmentation [20, 26], which trains a deep neural network from a set of training samples. However, their performance heavily depends on a large number of high-quality training samples in which each pixel has a label. In practice, the labeling process is time-consuming [50] and need many human efforts, especially in domain-specific applications such as medical images and seismic data. To relax the supervision requirements, some recent works utilize weakly or semi-supervised image segmentation, e.g., relaxations of the pixel-level annotations to image level [32] or bound box level [15]. These methods still need many training pairs to achieve good performance such that sufficiently many complex scenes are covered. Therefore, further relaxing the supervision requirements to