Journal of Image and Signal Processing 图像与信号处理, 2025, 14(2), 213-223
Published Online April 2025 in Hans. https://www.hanspub.org/journal/jisp
https://doi.org/10.12677/jisp.2025.142020

Hans汉斯

基于编码器－解码器架构和Potts模型的图像分割方法

张同德，潘振宽*，魏伟波，王烨然

青岛大学计算机科学技术学院，山东 青岛

收稿日期：2025年2月28日；录用日期：2025年3月20日；发布日期：2025年3月31日

摘　要

深度学习驱动的图像分割在医学影像和自动驾驶等领域成效显著，但其黑箱决策机制导致模型选择与超参数调整缺乏理论指导，依赖大数据和高算力支撑。相较之下，基于变分模型的方法虽多受限于局部特征提取，易忽略全局上下文关联，但其通过融合全局统计规律与局部平滑约束的特性，在数学可解释性和抗噪声伪影方面展现优势。因此，本文提出了一种基于Potts模型展开的与U-Net相似的架构，旨在提升图像分割的准确性和鲁棒性。与传统U-Net不同，本文在下采样和上采样过程中引入了基于Potts模型的正则化块，以增强分割过程中的区域一致性和边缘保留能力。通过HQS (半二次分裂)方法求解Potts模型，并结合FoE正则化项，使用可训练的离散余弦变换(DCT)-高斯卷积实现了梯度算子的学习，激活函数采用软阈值公式(STF)。此外，为了捕获全局上下文信息并处理远距离依赖，在网络的最底层加入了Transformer结构，进一步改善分割效果。实验结果表明，本文提出的模型在少量参数和数据集上能够有效学习特征，提高分割精度。本研究为图像分割任务提供了新的视角，展示了结合深度神经网络与传统变分模型架构的广阔潜力。

关键词

Potts模型，深度学习，变分网络，图像分割

Image Segmentation Method Based on Encoder-Decoder Architecture and Potts Model

Tongde Zhang, Zhenkuan Pan*, Weibo Wei, Yeran Wang

College of Computer Science & Technology, Qingdao University, Qingdao Shandon

Received: Feb. 28th, 2025; accepted: Mar. 20th, 2025; published: Mar. 31st, 2025

文章引用: 张同德, 潘振宽, 魏伟波, 王烨然. 基于编码器–解码器架构和 Potts 模型的图像分割方法[J]. 图像与信号处理, 2025, 14(2): 213-223. DOI: 10.12677/jisp.2025.142020

Abstract

Deep learning-driven image segmentation has demonstrated significant efficacy in fields such as medical imaging and autonomous driving. However, its black-box decision-making mechanisms lead to a lack of theoretical guidance for model selection and hyperparameter tuning, with heavy reliance on large datasets and high computational resources. In contrast, variational model-based methods, though often limited by local feature extraction and neglect of global contextual relationships, exhibit advantages in mathematical interpretability and noise/artifact resistance through their integration of global statistical patterns and local smoothness constraints. This paper proposes a U-Net-inspired architecture based on the Potts model unfolding framework to enhance segmentation accuracy and robustness. Unlike traditional U-Net, our method introduces Potts model-derived regularization blocks during downsampling and upsampling to strengthen region consistency and edge preservation capabilities. The Potts model is solved via the Half-Quadratic Splitting (HQS) method, combined with a Fields of Experts (FoE) regularization term. Trainable Discrete Cosine Transform (DCT)-Gaussian convolutions are employed to learn gradient operators, with activation functions adopting the Soft Thresholding Formula (STF). Additionally, a Transformer structure is integrated at the network's deepest layer to capture global contextual information and address long-range dependencies, further refining segmentation performance. Experimental results demonstrate that our model effectively learns features with limited parameters and datasets while improving segmentation precision. This study offers a novel perspective for image segmentation tasks, highlighting the vast potential of hybrid architectures that combine deep neural networks with classical variational models.

Keywords

Potts Model, Deep Learning, Variational Network, Image Segmentation

1. 引言

图像分割是计算机视觉中的一个核心任务，其目的是将图像划分为具有相似性质的多个区域，以便于后续分析或处理。近年来，随着技术的发展，图像分割方法经历了显著的变革。从最早的基于变分模型的方法，到后来的深度学习模型，这一领域逐步演变，展现出更加丰富和高效的分割策略。

在过去的几十年中，变分模型不仅建模过程简单、实现简单、算法稳定，并且有着强大的数学基础和理论支撑，成为图像分割的重要工具。这类方法通过最小化一个能量函数，将分割问题转换为优化问题，并利用全局信息和局部平滑性去控制分割结果。图像分割方法可以分为基于阈值的分割方法[1]、基于区域的分割方法[2]、基于边缘的分割方法[3]等。

1988 年，Osher 和 Sethian[4]提出了水平集方法(Level Set Method)，该方法用于跟踪曲线和曲面的演化，具有处理变形和分裂的能力。在之后的研究中，Chan 和 Vese 进一步将这一概念结合到图像分割中，提出了 Chan-Vese 模型[5]。该模型通过最小化特定的能量函数，基于变分方法实现了图像的两相分割，允许处理灰度图和具有纹理的图像。Chan-Vese 模型特别强调了图像区域的均匀性，改进了传统的阈值分割方法，使其具有更好的鲁棒性和灵活性。后来 Chan 和 Samson 又将这些技术扩展到多相图像分割问题

[6]。2010 年，潘振宽等人又提出了一种基于 TV 模型的多相图像分割变分水平集方法[7]。变分方法具有透明和可解释的数学公式的优势，并且不需要太多的训练数据，这对于医学图像分析任务非常重要。但是，它们仍然存在局限性，例如对初始化设置的敏感性，需要对不同的数据进行手动参数调整，在处理复杂背景、噪声和物体边界时面临挑战，尤其是在多相分割任务中。

随着深度学习的兴起，尤其是卷积神经网络(CNN)的成功应用，图像分割领域发生了颠覆性变革。2015 年，Long 等人提出了全卷积网络(FCN) [8]，作者将传统的卷积神经网络(CNN)架构改进为能够直接输出像素级别的分割结果，标志着深度学习在图像分割领域的应用取得了重大突破。同年，Ronneberger 等人又提出了一种专门为生物医学图像分割设计的深度卷积网络架构——U-Net [9]，作者提出了一种对称的编码-解码结构，其中的跳跃连接有助于保留高分辨率特征，使得网络能够在像素级进行更准确的分割。U-Net 因在医学成像中表现优异而广受欢迎，成为许多后续研究的基准模型。

尽管深度学习模型在图像分割方面展现了优越的性能，但依然存在一些不足之处。例如，深度网络的训练需要大量标注数据及计算资源，且其决策过程常常缺乏可解释性。新提出的 SAM (Segment Anything Model) [10]也不例外，它自然图像领域的表现优异，但在医学图像领域，由于患者人口统计学差异和极端病理情况的存在，SAM 的表现并不是那么优异。此外，其海量的参数，在算力不足的情况下(如移动设备)进行部署变得具有挑战性。这使得研究者们逐渐认识到，结合传统的变分模型和现代深度学习的优势，或许能够在分割质量和模型解释性之间找到一个更好的平衡点。

近年来，学者们探索了深度学习框架和变分模型之间的协同作用。Chen 等人[11]将全卷积网络(FCN)与条件随机场(CRF)结合，提出了一种新的分割方法，显示了深度学习和图模型之间的协同作用。Y. Chen [12]提出了一种基于变分推理的图像分割方法，通过结合深度学习生成对抗网络(GAN)和变分模型，显著提高图像分割的鲁棒性和准确性。Tai 等人[13]将 Potts 模型与多重网格方法结合，通过对 Potts 模型的离散化和分析，提供了一种数学上合理的解释，阐明了编码器-解码器网络的工作机制。尽管上述方法将传统的变分方法集成到深度神经网络中，但对于网络整体训练和底层行为的清晰数学解释仍然难以把握。探索基于经典变分模型的可学习网络工作，可以有效减少手动参数调整和对大量数据集的依赖，这是一项具有创新性的重要尝试。在图像恢复领域，Chen 和 Pock 将非线性反应扩散方程与深度学习模型结合，提出了可训练非线性反应扩散(TNRD) [14]使其能够自动学习适应于不同图像恢复任务的参数。该模型使用径向基函数(RBF)和卷积核的组合来创建可训练的参数激活函数，为我们的研究提供了新的见解。2017 年，Vaswani 等人首次提出了 Transformer 架构[15]，推动了自然语言处理(NLP)的突破，使得基于深度学习的各类应用得以快速发展。后来，Carion 等人提出 DETR (DEtection TRansformer)模型[16]，尽管 DETR 的主要目标是实现目标检测，但其框架和自注意力机制也为图像分割任务提供了新的思路。

在多相分割的研究中，提升成像质量和目标检测能力的需求使得对多类别的分割性能要求日益增强。为了有效应对这一挑战，本研究提出了一种结合 Potts 模型[17]的类似 U-Net 架构的新的图像分割方法，该方法利用正则化模块将正则化项的迭代过程嵌入到通道数较少的浅层网络中。并在网络最深层引入 TransFormer 来捕获全局上下文信息并处理远距离依赖，旨在提高多相分割的准确性与鲁棒性。

2. 相关研究基础

多相图像分割是根据不同的图像特征将图像分为多个互相毗邻、互不重叠的区域，其数学描述为 $\Omega=\bigcup_{i=1}^{n}\Omega_i$，$\Omega_i\cap_{i\neq j}\Omega_j=\varnothing$，其中 Ω 为待分割区域，$\Omega_i\left(i=1,2,\cdots,n\right)$ 为分割出的多个子区域。

2.1. Potts 模型

Potts 模型最初源于统计物理。在图像分割的背景下，Potts 模型将图像的每个像素视为一个节点，并

将其划分为不同的类别(状态)。相邻像素的分类状态影响这两个像素之间的相互关系和能量。Potts 模型通过定义一个能量函数来表示图像分割的目标，其能量函数可以表示为：

$$\min_{c,\varphi\in\{0,1\}} E(c,\varphi)=\sum_{i=1}^{n}\alpha_i\int\left(f(x)-c_i\right)^2\varphi_i\mathrm{d}x+\gamma\sum_{i=1}^{n}\left|\nabla\varphi_i\right|\mathrm{d}x \tag{1}$$

它由数据项和总变分(Total Variation, TV)正则化项组成。其中 $f(x)$ 表示输入图像； c_i 表示每个分割区域的均值；φ_i 是指示函数，用于指示像素 x 是否属于区域 i。使用 Simplex 法对 φ 进行约束，使 $\sum_{i=1}^{n}\varphi_i=1$；$\alpha_i$ 和 γ 是惩罚参数。

方程(1)的求解方法引起了许多学者的关注，Zach 等人[18]将标记函数松弛为非凸形式，通过交替迭代优化获得解决方案。Bae 等人[19]提出了一个原始对偶模型，并使用投影梯度算法解决。

2.2. HQS 方法

Half Quadratic Splitting (HQS，半二次分裂法)是一种有效的优化算法，主要用于解决涉及非凸和非平滑函数的数值优化问题。引入辅助变量 $\omega_i=\nabla\varphi_i$ 和惩罚参数 β_i，并定义数据项为 $g_i(x)=\left(f(x)-c_i\right)^2$，则能量函数变化为：

$$E(c,\varphi,\omega)=\sum_{i=1}^{n}\alpha_i\int g_i(x)\varphi_i\mathrm{d}x+\sum_{i=1}^{n}\gamma_i\int\left|\omega_i\right|\mathrm{d}x+\frac{\beta_i}{2}\sum_{i=1}^{n}\int\left(\omega_i-\nabla\varphi_i\right)^2\mathrm{d}x \tag{2}$$

通过采用交替优化方法进行进一步求解，得到了 ω 的解析解：

$$\omega_i^{l+1}=\max\left(\left|\nabla\varphi_i^l\right|-\frac{\gamma_i}{\beta_i},0\right)\frac{\left|\nabla\varphi_i^l\right|}{\nabla\varphi_i^l} \tag{3}$$

其中，l 表示当前迭代的步数

3. 研究方法

3.1. FoE (Field of Experts)正则化[20]

FoE 正则化是一种有效的图像正则化方法，它采用非线性和非凸的形式，能够更好地捕捉图像中的复杂结构和纹理。并且，FoE 模型的参数可以通过训练数据进行学习。一般情况，FoE 正则化项通常可以表示为：

$$R_{FoE}(x)=\sum_{i=1}^{n}\varphi\left(\sum_{j\in N_i}K_{i,j}x_j\right) \tag{4}$$

其中，x 是图像特征，n 是专家场的数量，N_i 表示第 i 个专家场的邻域，$K_{i,j}$ 表示第 i 个专家场的卷积核，φ 是一个非线性激活函数。据此，可以得出带有 FoE 正则化项的 Potts 模型：

$$\min_{c,\varphi\in\{0,1\}} E(c,\varphi)=\sum_{i=1}^{n}\alpha_i\int g_i(x)\varphi_i\mathrm{d}x+\sum_{i=1}^{n}\sum_{j=1}^{m}\gamma_i\int\left|K_j\varphi_i\right|\mathrm{d}x \tag{5}$$

这里 γ_i 表示权重，基于变分法和梯度下降策略进行求解，结果如下：

$$\frac{\partial\varphi}{\partial t}=-\alpha_i g_i(x)-\sum_{j=1}^{m}\beta_{ji}\tilde{K}_j\left(K_i\varphi_i^l-\omega_j^l\right) \tag{6}$$

$$\omega_i^{l+1}=\max\left(\left|K_j\varphi_i^l\right|-\frac{\gamma_i}{\beta_i},0\right)\frac{K_j\varphi_i^l}{\left|K_j\varphi_i^l\right|} \tag{7}$$

其中，$\varphi_i^l \in [0,1]$，$\tilde{K}_j$ 是 K_i 的共轭，方程(7)是 φ_i 的软阈值公式，是一个解析解，根据其性质，可以将其转化为以下形式：

$$\omega_{ji}^{l+1} = \mathrm{ReLU}\left(K_j\varphi_i^l - \frac{\gamma_i}{\beta_{ji}}\right) - \mathrm{ReLU}\left(-\left(K_j\varphi_i^l + \frac{\gamma_i}{\beta_{ji}}\right)\right) \tag{8}$$

这将作为激活函数用在网络中。

3.2. Vision Transformer 模块集成

Vision Transformer (ViT)是一种将 Transformer 架构应用于计算机视觉任务的模型，最初由 Dosovitskiy [21]等人在 2020 年提出。该方法通过将图像划分为小块，并将这些小块作为序列输入给 Transformer，从而实现了对图像的处理和分类。

下采样路径处理完特征图后，嵌入一个 ViT 模块。具体而言，在下采样部分生成的 $H\times W$ 特征图分块成 $P\times P$ 个小块，生成块的数量 $N = \frac{H\times W}{P^2}$。然后每个小块被展平为一个一维向量，并通过一个线性投影层(线性变换)将其映射到 D 维特征空间。

$$\mathrm{PatchEmbedding}: Z = \mathrm{Proj}\left(\mathrm{Flatten}\left(\mathrm{Patch}\left(I\right)\right)\right) \tag{9}$$

其中，I 是输入图像，$\mathrm{Patch}\left(I\right)$ 表示对图像进行分块操作，Flatten 表示将每个小块展平，Proj 是线性变换。

为了保持位置信息，这一序列还将加入位置编码。Transformer 模块由多头自注意力机制和前馈神经网络组成，能够在全局范围内捕捉特征之间的复杂关系，其输出可表示为：

$$\mathrm{Attention}\left(Q,K,V\right) = \mathrm{softmax}\left(\frac{QK^{\mathrm{T}}}{\sqrt{d_k}}\right)V \tag{10}$$

其中，Q 是查询(query)，K 是键(key)，V 是值(value)，d_k 是键的维度。经过 ViT 处理后的特征图准备进行后续的上采样操作。

3.3. 网络设计

根据以上内容，将原本缺乏学习能力的 Potts 模型拓展为可训练的网络架构，为了学习复杂特征和捕捉长距离依赖关系，本文设计了一个基于编码器-解码器架构的网络，在下采样和上采样过程中，在通道数较少的浅层都加入了正则化模块，并在最深层加入了 ViT 模块，如图 1 所示。

该类似 U-Net 框架的下采样阶段用公式(5)中的数据项实现，浅层的正则化块作为正则化项的实现，对应的偏微分方程更新如下：

$$\frac{\partial\varphi}{\partial t} = U\left(f\left(x\right)\right) - \sum_{j=1}^{m}\beta_{ji}\tilde{K}_j\left(K_j\varphi_i^l - \omega_j^l\right) \tag{11}$$

U 表示网络中 U-Net 结构内使用的下采样过程和上采样过程。在正则化模块中，将微分算子 K 转换为卷积计算。受 TNRD 的启发，为了显著减少参数数量并确保卷积核的零均值特性，本文考虑将卷积核 K 设置为以下 DCT 基的线性组合：

$$K_j = \sum_{r=1}^{d}\frac{w_{j,r}}{\left\|w_j\right\|}b_r \tag{12}$$

其中，$r = 1,2,\cdots,d$。$w_j = \left(w_{j1},w_{j2},\cdots,w_{jd}\right)$ 表示可训练的参数。为了减少对噪声或不相关局部特征的敏感程度，将一个二维高斯核函数引入卷积核，作用于 φ_i：

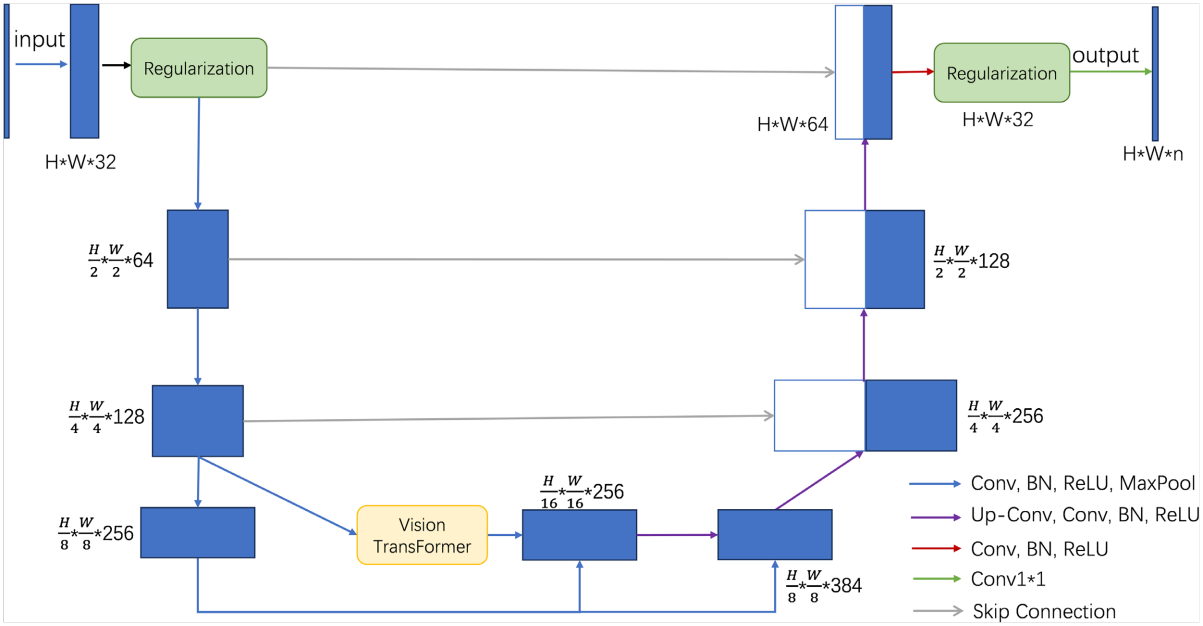

Figure 1. The network structure
图 1. 本文提出的网络结构

$$G(x,y)=\frac{1}{2\pi\sigma^2}\exp\left(-\frac{x^2+y^2}{2\sigma^2}\right) \tag{13}$$

其中，σ 表示高斯函数的标准差，是一个可训练的参数。最后将正则化模块设置为离散余弦变换基和高斯核函数矩阵的乘积：

$$K_j=G(\sigma)\times\sum_{r=1}^{d}\frac{w_{j,r}}{\|w_j\|}b_r \tag{14}$$

对于共轭差分算子 $\tilde{K}_j$，将前一个卷积核旋转 180 度，并应用反卷积来计算共轭。

在这两个卷积之间，将批量归一化(BN)和等式(8)中的平滑 STF 作为激活层，最终的正则化模块如图 2 所示。

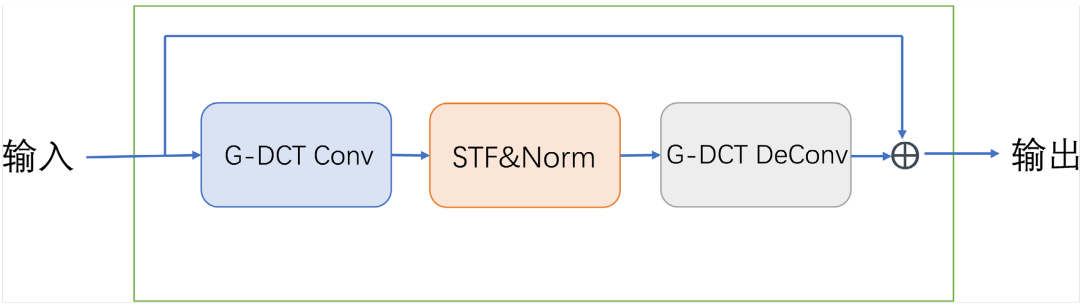

Figure 2. Regularization module structure
图 2. 正则化模块结构

图 1 网络将原始图像作为输入，并在输入图像上初始化特征函数 φ_i，以解决传统变分模型的初始化敏感性问题。最后通过 Softmax 和 Simplex 约束，得出每个像素的分类结果，从而生成最终的图像分

割掩膜。

通过反向传播训练和更新网络参数集 $\theta^l=\left\{K_j^l,\alpha^l,\beta^l,\gamma^l,\sigma^l\right\}$，使用 Adam 优化器[22]最小化以下交叉熵函数损失：

$$L=-\frac{1}{N}\sum_{i=1}^{N}\sum_{c=1}^{C}y_{i,c}\log\left(p_{i,c}\right) \tag{15}$$

其中，N 是样本的数量，C 是类别的数量，$y_{i,c}$ 表示样本 i 在类别 c 下的真实标签，$p_{i,c}$ 表示样本 i 在类别 c 下的预测概率。

4. 实验分析

4.1. 实验环境

实验是在配备 Intel Core i7-12700 处理器、NVIDIA GeForce GTX 2080Ti 显卡的主机上进行的。初始学习率设置为 0.002，epoch 设置为 200。将该模型与流行的图像分割网络进行了比较，包括 U-Net、U-Net++ [23]、FPN [24]、DeepLabV3+ [25]。

在实验中，为了全面评估模型的分割性能，本实验使用了多种指标：

1) Accuracy (准确性)：衡量模型预测正确的比例，即正确预测的像素数占总像素数的百分比。

2) DSC (Dice Similarity Coefficient，Dice 相似系数)：用于测量两个集合的相似度，特别适合评估分割结果与真实标注的重合程度。其值在 0 到 1 之间，1 表示完全一致。公式为：

$$\mathrm{DSC}=\frac{2\left|A\cap B\right|}{\left|A\right|+\left|B\right|} \tag{16}$$

其中，$\left|A\cap B\right|$ 是预测与真实标注的交集面积。

3) MIoU (Mean Intersection over Union，平均交并比)：评估预测结果与真实值的重叠程度，是在所有类别上的 IoU 的平均值。IoU 计算公式为：

$$\mathrm{IoU}=\frac{\left|A\cap B\right|}{\left|A\cup B\right|} \tag{17}$$

这些指标用百分比表示，可以直观地反映模型的分割性能。通过结合这些方法，能够更全面地理解模型的优劣。

4.2. 数据集

本文选择了两个数据集：

1) Synapse 医学图像数据集，它包含 2211 张人体 CT 扫描图像，分辨率为 512 × 512，本实验选取了其中 1280 张人体腹部的 CT 扫描图像，其中 1000 张作为训练集，280 张作为测试集。

2) CamVid (Cambridge Driving Labeled Video Database)，它是由剑桥大学的计算机视觉组创建，是一个广泛用于深度学习和计算机视觉领域的语义分割数据集，特别是在自动驾驶和道路场景理解方面。它包含 701 张分辨率为 360 × 480 图像。本实验使用 366 张作为训练集，233 张作为测试集，102 张作为验证集。

4.3. 实验结果

首先是在 Synapse 医学图像数据集上的分割，本实验专注于肝脏(分割结果用绿色表示)、脾脏(分割结果用黄色表示)的分割。在实验结果中，选取了五组分割效果展示，如图 3 所示。

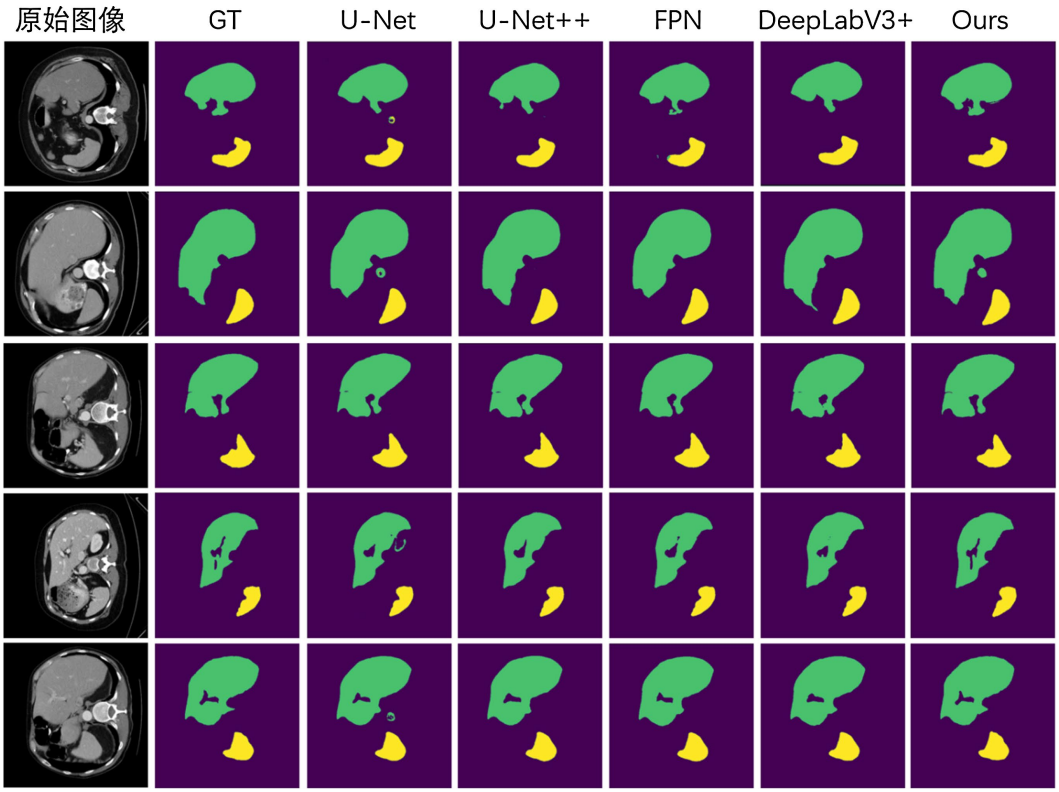

Figure 3. Comparative experiments on the Synapse dataset
图 3. Synapse 医学图像数据集上的对比实验

可以看出，相较于同为解码器-编码器结构的 U-Net、U-Net++来说，本文带有正则化块的模型更能分割出清晰的边界，不会遗漏边缘特征不明显的区域。较之使用空洞卷积来获取更多的上下文信息的 DeepLabV3+来说，使用 Vision TransFormer 的模型更能从边缘特征不明显的图像中得到正确的边界。

表 1 展示了该模型的性能指标，可以看出该模型的指标全面优于 U-Net、U-Net++、FPN。并且，该模型的参数量仅为 DeepLabV3+的 8%，却达到了与其难分伯仲的效果，表明该模型在特征提取和分割精度上具有更好的平衡性。

Table 1. Comparison of segmentation results on the Synapse dataset
表 1. Synapse 医学图像数据集上的分割结果对比

Model	Accuracy	DSC	MIoU	Number of Parameters
Ours	**97.11**	**95.94**	96.56	**3.9M**
U-Net	95.87	93.26	94.26	31M
U-Net++	96.44	93.79	94.77	31.5M
FPN	96.66	94.55	94.63	29M
DeepLabV3+	97.06	95.88	**96.61**	49.4M

为了进一步评估模型的效果，本文选择了较为复杂的街景数据集 **CamVid** 来进行试验，并选取了一些试验的对比结果，如图 4 所示。

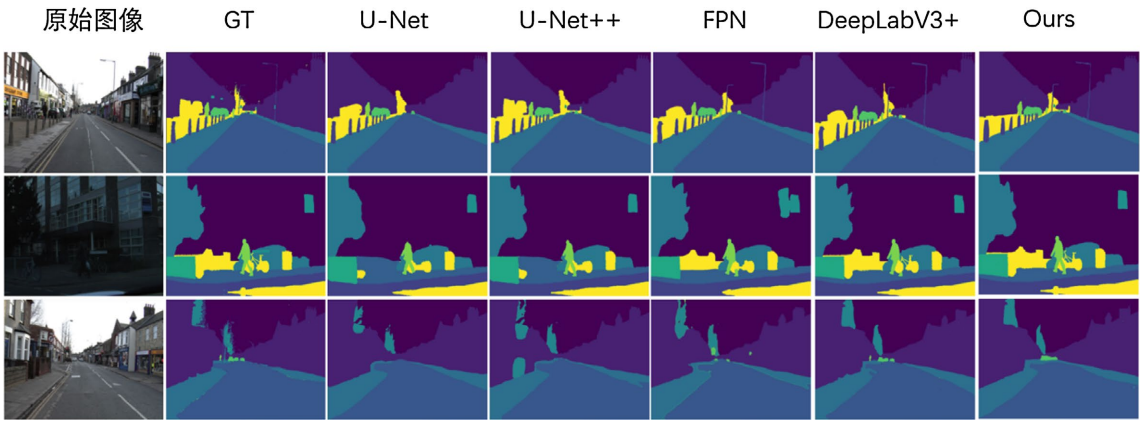

Figure 4. Comparative experiment on CamVid dataset
图 4. 在 CamVid 数据集上的对比实验

可以看出，该模型具有高分辨率的特征提取能力，能够有效地捕捉图像中的细节特征，例如树冠、电线杆、自行车等。同时可以在保留细节的同时，生成更加平滑的边缘，在细节较多的图像中可以实现较好的分割效果。

Table 2. Comparison of segmentation results on Camvid dataset
表 2. Camvid 数据集上的分割结果对比

Model	Accuracy	DSC	MioU	Number of Parameters
Ours	**97.42**	**95.06**	**96.98**	**3.9M**
U-Net	95.85	93.57	94.42	31M
U-Net++	96.09	93.90	94.83	31.5M
FPN	96.61	94.35	96.17	29M
DeepLabV3+	97.11	94.76	96.75	49.4M

表 2 展示了该模型在 **CamVid** 的表现，实验结果表明，该模型在精度、DSC 和 MIoU 指标上均优于其他模型。由于路面、天空、建筑占据图像较大区域，使得这些模型的整体指标都较高且表现相近，但是还可以看出，该模型整体性能优于参与比较的这些模型。

4.4. 消融实验

为了探究网络中正则化模块和 ViT 的有效性，本文在 Synapse 医学图像数据集上进行了消融实验，首先是正则化模块的有效性验证，结果如表 3 所示。可以看出，正则化模块的引入，确实使模型的性能有了显著的提高。

Table 3. An ablation study to verify the effectiveness of the regularization module
表 3. 验证正则化块有效性的消融实验

	Accuracy	DSC	MioU
有正则化块	**97.11**	**95.94**	**96.56**
无正则化块	96.87	95.13	95.72

接下来是 ViT 的有效性验证结果如表 4 所示。可以看出，ViT 模块的引入使得该模型从各个指标上

都有了显著的提升。

Table 4. An ablation study to verify the effectiveness of the ViT module
表 4. 验证正则化块有效性的消融实验

	Accuracy	DSC	MioU
有 ViT	**97.11**	**95.94**	**96.56**
无 ViT	96.43	94.87	95.36

总体实验结果表明，所提出的网络在 Synapse 医学图像和 CamVid 街景图像上均取得了较好的分割结果。在与当前主流方法进行对比时，模型在准确性、Dice 相似度系数(DSC)和平均交并比(MIoU)等关键指标上均表现出色，表明本文提出的模型能够有效提升细小物体的分割能力及整体性能。

5. 总结

本论文提出了一种新型的基于编码器-解码器的网络架构，该架构将传统变分方法与深度学习相结合，以提升图像分割和特征提取的效果。该模型有效解决了变分模型的初始化敏感性问题，以及深度学习模型对大量数据和可解释性要求的挑战。通过引入可学习的正则化块，我们能够有效捕捉图像中的全局特征和细节信息，同时利用 ViT 的自注意力机制优化特征表示，增强模型对复杂场景的理解能力。

此外，本研究的结果揭示了将传统变分方法与现代深度学习技术结合的潜力，未来的研究将集中在如何进一步优化变分模型的参数调整机制以及提升模型在更复杂场景中的适用性，期待能在不断变化的视觉环境中探索更多应用场景。

参考文献

[1] Otsu, N. (1979) A Threshold Selection Method from Gray-Level Histograms. *IEEE Transactions on Systems*, *Man*, *and Cybernetics*, **9**, 62-66. https://doi.org/10.1109/tsmc.1979.4310076

[2] Adams, R. and Bischof, L. (1994) Seeded Region Growing. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **16**, 641-647. https://doi.org/10.1109/34.295913

[3] Canny, J. (1986) A Computational Approach to Edge Detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **8**, 679-698. https://doi.org/10.1109/tpami.1986.4767851

[4] Osher, S. and Sethian, J.A. (1988) Fronts Propagating with Curvature-Dependent Speed: Algorithms Based on Hamilton-Jacobi Formulations. *Journal of Computational Physics*, **79**, 12-49. https://doi.org/10.1016/0021-9991(88)90002-2

[5] Chan, T.F. and Vese, L.A. (2001) Active Contours without Edges. *IEEE Transactions on Image Processing*, **10**, 266-277. https://doi.org/10.1109/83.902291

[6] Samson, C., Blanc-Féraud, L., Aubert, G. and Zerubia, J. (2000) A Level Set Model for Image Classification. *International Journal of Computer Vision*, **40**, 187-197. https://doi.org/10.1023/a:1008183109594

[7] 李忠伟, 潘振宽, 倪明玖. 基于 TV 模型的多相图像分割变分水平集方法[C]//第五届图像图形技术与应用学术会议论文集. 北京: 北京图像图形学学会, 2010: 49-56.

[8] Long, J., Shelhamer, E. and Darrell, T. (2015) Fully Convolutional Networks for Semantic Segmentation. 2015 *IEEE Conference on Computer Vision and Pattern Recognition* (*CVPR*), Boston, 7-12 June 2015, 3431-3440. https://doi.org/10.1109/cvpr.2015.7298965

[9] Ronneberger, O., Fischer, P. and Brox, T. (2015) U-Net: Convolutional Networks for Biomedical Image Segmentation. *Medical Image Computing and Computer-Assisted Intervention—MICCAI* 2015, Munich, 5-9 October 2015, 234-241. https://doi.org/10.1007/978-3-319-24574-4_28

[10] Kirillov, A., Wu, Y., He, K. and Girshick, R. (2023) Segment Anything. *Proceedings of the IEEE/CVF International Conference on Computer Vision*, Paris, 2-6 October 2023, 4015-4026.

[11] Chen, X., Papandreou, G. and Schroff, F. (2016) Semantic Image Segmentation with Deep Convolutional Nets and Fully Connected Conditional Random Fields. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (*CVPR*), Las Vegas, 27-30 June 2016, 2437-2446.

[12] Chen, Y., Qi, H. and Dee, D. (2020) Deep Variational Image Segmentation via a Conditional Generative Model. *Proceedings of the AAAI Conference on Artificial Intelligence*, **34**, 10843-10850.

[13] Tai, X., Liu, H. and Chan, R. (2024) PottsMGnet: A Mathematical Explanation of Encoder-Decoder Based Neural Networks. *SIAM Journal on Imaging Sciences*, **17**, 540-594. https://doi.org/10.1137/23m1586355

[14] Chen, Y. and Pock, T. (2017) Trainable Nonlinear Reaction Diffusion: A Flexible Framework for Fast and Effective Image Restoration. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **39**, 1256-1272. https://doi.org/10.1109/tpami.2016.2596743

[15] Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, L., Polosukhin, M. and Tsipras, S. (2017) Attention Is All You Need. *Proceedings of the* 31*st International Conference on Neural Information Processing Systems*, Long Beach, 4-9 December 2017, 6000-6010.

[16] Carion, N., Massa, F., Synnaeve, G., Usunier, N., Kirillov, A. and Zagoruyko, S. (2020) End-to-End Object Detection with Transformers. 16*th European Conference on Computer Vision*, Glasgow, 23-28 August 2020, 213-229. https://doi.org/10.1007/978-3-030-58452-8_13

[17] Shotton, J.B., Winn, J.M., Rother, C. and Torr, P.H. (2006) TextonBoost for Image Segmentation and Recognition. *Proceedings of the European Conference on Computer Vision* (*ECCV*), Graz, 7-13 May 2006, 1-15.

[18] Zach, C., Gallup, D., Frahm, J.M. and Niethammer, M. (2008) Fast Global Labeling for Real-Time Stereo Using Multiple Plane Sweeps. 13*th International Fall Workshop on Vision*, *Modeling*, *and Visualization*, *VMV* 2008, Konstanz, 8-10 October 2008, 243-252.

[19] Bae, E., Yuan, J. and Tai, X. (2010) Global Minimization for Continuous Multiphase Partitioning Problems Using a Dual Approach. *International Journal of Computer Vision*, **92**, 112-129. https://doi.org/10.1007/s11263-010-0406-y

[20] Roth, S. and Black, M.J. (2009) Fields of Experts. *International Journal of Computer Vision*, **82**, 205-229. https://doi.org/10.1007/s11263-008-0197-6

[21] Dosovitskiy, A., Beyer, D., Kolesnikov, A., Zhai, X. and Hoffmann, T. (2021) An Image Is Worth 16 x 16 Words: Transformers for Image Recognition at Scale. *Proceedings of the International Conference on Learning Representations* (*ICLR*), 3-7 May 2021.

[22] Kingma, D.P. and Ba, J. (2015) Adam: A Method for Stochastic Optimization. *Proceedings of the* 3*rd International Conference on Learning Representations* (*ICLR*), San Diego, 7-9 May 2015, 1-15.

[23] Zhou, Z., Rahman Siddiquee, M.M., Tajbakhsh, N. and Liang, J. (2018) Unet++: A Nested U-Net Architecture for Medical Image Segmentation. In: Stoyanov, D., *et al*., Eds., *Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support*, Springer International Publishing, 3-11. https://doi.org/10.1007/978-3-030-00889-5_1

[24] Lin, T., Dollar, P., Girshick, R., He, K., Hariharan, B. and Belongie, S. (2017) Feature Pyramid Networks for Object Detection. 2017 *IEEE Conference on Computer Vision and Pattern Recognition* (*CVPR*), Honolulu, 21-26 July 2017, 936-944. https://doi.org/10.1109/cvpr.2017.106

[25] Chen, L. C., Papandreou, G., Schroff, F., Adam, H. (2017) Rethinking Atrous Convolution for Semantic Image Segmentation. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (*CVPR*), Honolulu, 21-26 July 2017, 1-10.