

电商智能质检与营销素材处理的高效解决方案

——基于半监督语义分割的数字图像技术应用

王自祥

上海理工大学机械工程学院, 上海

收稿日期: 2026年2月5日; 录用日期: 2026年2月14日; 发布日期: 2026年3月11日

摘要

在电商智能运营、网络素材营销及互联网图像处理中, 商品缺陷检测、物流包裹分拣、营销素材标注等大量图像数据的语义分析依赖人工标注, 存在运营成本高、效率低的问题; 且单一阈值伪标签方法易浪费有效数据, 错误累积还会影响决策精度。为破解电商领域图像语义标注的成本难题与智能分析精度瓶颈, 本文提出数字图像技术赋能的电商图像智能处理半监督语义分割方法MvpMatch, 将分层阈值策略与历史模型集成预测机制深度适配电商场景。分层阈值通过多级置信区间与差异化损失权重, 高效利用海量商品展示图、物流流转图等电商无标签数据; 历史模型集成预测通过多视角聚合降低伪标签错误累积, 提升分析可靠性。基于电商营销素材Marketing数据集的实验表明, 该方法在不同标签比例下均显著优于基线模型UniMatch V2, 少标签场景下mIoU最高提升1.9个百分点。研究结果可直接应用于电商智能质检、自动化分拣、营销素材批量处理等核心环节, 助力降低企业信息化成本、提升网络营销精度与运营绩效, 为数字技术与电商业深度融合提供技术支撑。

关键词

数字图像技术, 电商智能运营, 网络营销, 分层阈值, 半监督语义分割

An Efficient Solution for E-Commerce Intelligent Quality Inspection and Marketing Material Processing

—Application of Digital Image Technology Based on Semi-Supervised Semantic Segmentation

Zixiang Wang

Abstract

In intelligent e-commerce operations, online marketing content, and web image processing, semantic analysis of large-scale image data—including product defect detection, logistics parcel sorting, and marketing asset labeling—relies heavily on manual annotation, resulting in high operational costs and low efficiency. Moreover, single-threshold pseudo-label methods tend to waste valid data, and error accumulation compromises decision accuracy. To address the cost challenges of image semantic annotation and bottlenecks in intelligent analysis accuracy within e-commerce, this study proposes Mvp-Match, a digital image-empowered semi-supervised semantic segmentation method for intelligent e-commerce image processing. This method deeply adapts hierarchical thresholding strategies and temporal model ensemble prediction mechanisms to e-commerce scenarios. Hierarchical thresholding efficiently leverages massive unlabeled data—including product display images and logistics flow images—through tiered confidence thresholds and adaptive loss weighting. Temporal model ensemble prediction reduces pseudo-label error accumulation through multi-view aggregation, enhancing analysis reliability. Experiments on the e-commerce marketing asset Marketing-E dataset demonstrate that the proposed method significantly outperforms the baseline model UniMatch V2 across various labeling ratios, achieving mIoU improvements of up to 1.9 percentage points in low-label regimes. These findings can be directly applied to core e-commerce processes such as intelligent quality inspection, automated sorting, and batch processing of marketing assets, helping reduce operational IT costs, improve online marketing precision and operational performance, and providing technical support for deeper convergence of digital technologies with e-commerce operations.

Keywords

Digital Image Technology, E-Commerce Intelligent Operation, Online Marketing, Hierarchical Threshold, Semi-Supervised Semantic Segmentation

Copyright © 2026 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

1. 引言

在大数据与人工智能技术发展背景下，数字图像技术已成为推动电商数字化升级的核心支撑，其在商品缺陷检测、物流包裹分拣、营销素材处理等电商核心环节的应用，均依赖图像语义分割技术[1]实现像素级精准识别。传统全监督语义分割需大规模人工标注数据，而电商图像数据具有海量、多样、动态更新的特征，高成本的人工标注不仅耗时耗力，还易因标注延迟影响业务效率，成为制约电商信息化建设与智能化升级的关键瓶颈[2]。

半监督语义分割以“少量有标签数据 + 大量无标签数据”为学习模式，能有效降低人工标注依赖，是解决电商图像标注难题的优选路径。当前主流的一致性正则化方法在电商场景应用中仍存不足：一是单一阈值筛选伪标签，易丢弃模糊商品边缘、包裹轻微磨损区域等低置信度但有价值的像素，造成数据利用率偏低；二是模型自训练中伪标签错误易累积固化，导致图像分析结果偏差，影响后续业务决策的科学性。

针对上述问题, 本文提出融合分层阈值策略与历史模型集成预测的半监督语义分割方法 **MvpMatch**, 适配电商场景图像智能分析需求。该方法通过分层阈值充分挖掘电商海量无标签图像数据的价值, 借助历史模型集成预测降低伪标签错误累积的影响, 提升分析结果可靠性。本方法可直接应用于电商智能质检、物流自动化分拣、营销素材智能处理等环节, 为电商信息化建设提供低成本、高效率的技术解决方案, 助力电商图像分析环节的智能化升级。

2. 相关工作

2.1. 数字技术在电商中的应用现状

大数据与人工智能技术已广泛渗透电商全链路[3]: 在运营层面, 智能分拣、缺陷检测、库存可视化等技术提升供应链效率, 头部物流企业通过图像识别技术将分拣误差率降低至 0.3% 以下; 在网络营销层面, 图像语义分析助力营销素材优化、个性化推荐与虚拟试衣、试妆等场景落地, 某电商平台应用该技术后广告点击率提升 27%; 在企业信息化层面, 数据驱动的智能分析成为决策核心, 电商 ERP 系统已逐步集成图像智能处理模块。但现有电商图像分析方法多依赖全监督学习或简单半监督策略, 存在标注成本高、鲁棒性不足的问题, 难以适配电商数据“动态更新、场景多样”的特性——例如新品上架时缺乏标注数据导致智能质检系统无法快速适配, 直播电商的实时图像分析对模型精度与速度的双重要求难以满足。

2.2. 半监督语义分割在电商场景的研究进展

一致性正则化是半监督语义分割的主流范式, **FixMatch** [4]、**CPS** [5] 等方法都是使用这种基础框架, 但在电商场景的适配性不足: 单一阈值导致电商图像中复杂场景下的有效数据浪费(如反光商品的边缘像素), 错误伪标签累积影响商品检测、包裹分拣等关键环节的精度。**PrevMatch** [6] 等引入时间维度知识的方法, 为解决错误累积提供了思路, 但未针对电商图像“多场景、高噪声”的特点优化。近年来, 部分研究尝试将语义分割与电商场景结合: 一些互联网公司 2023 年 **Yang Lihe** 等提出的强弱一致性[7] 半监督语义分割方法用在电商商品缺陷检测上, 但未解决低置信度的数据利用问题; 还有一些图像处理公司将 2024 年 **Wang Haonan** 等提出的 **AllSpark** [8] 半监督语义分割方法应用于物流包裹识别, 但错误累积导致分拣误差率高于 1%, 难以满足实际运营需求。本文方法在 **UniMatch V2** [9] 基础上, 通过分层阈值与历史集成预测的双机制协同, 填补电商场景半监督语义分割“高利用率、低错误率”的技术空白。

3. 方法

本文提出的 **MvpMatch** 方法核心是: 将分层阈值策略与历史模型集成预测机制融入电商图像半监督语义分割, 适配电商运营中的精准性、高效性、低成本图像分析需求, 其整体框架如图 1 所示。

3.1. 预备知识

电商场景中, 半监督语义分割涉及两类数据: 带标签电商图像数据集 $D_l = \{x_i^l, y_i^l\}$, 如标注好的商品缺陷图、包裹特征图、营销素材标注图, 还有无标签电商图像数据集 $D_u = \{x_i^u\}$, 如海量未标注商品图、物流图、直播截图, 其中 x_i 为第 i 张图像, y_i 是为其对应的真实语义分割掩码, 且 D_u 规模远大于 D_l (通常比例为 10:1 至 100:1)。模型通过少量有标签数据初始化, 利用无标签数据挖掘语义特征, 实现电商图像的精准分割。

而常用的方法——基于一致性正则化的核心逻辑是: 对无标签电商图像施加强增强(裁剪、水平翻转等, 适配商品图标准化需求)与弱增强(颜色扰动、高斯模糊, 适配不同光照下的物流图、直播图), 在弱增强图像上生成伪标签, 在强增强图像上训练, 约束模型预测一致性, 总损失函数由有标签损失 L_x 与无标签损失 L_u 构成。在本篇文章的后续, 我们的基础公式的损失将沿用 **FixMatch** 损失逻辑, 并在 **UniMatch**

V2 的框架进行改进，针对电商图像特性做出优化参数配置，以此来适配电商场景和数据分布。

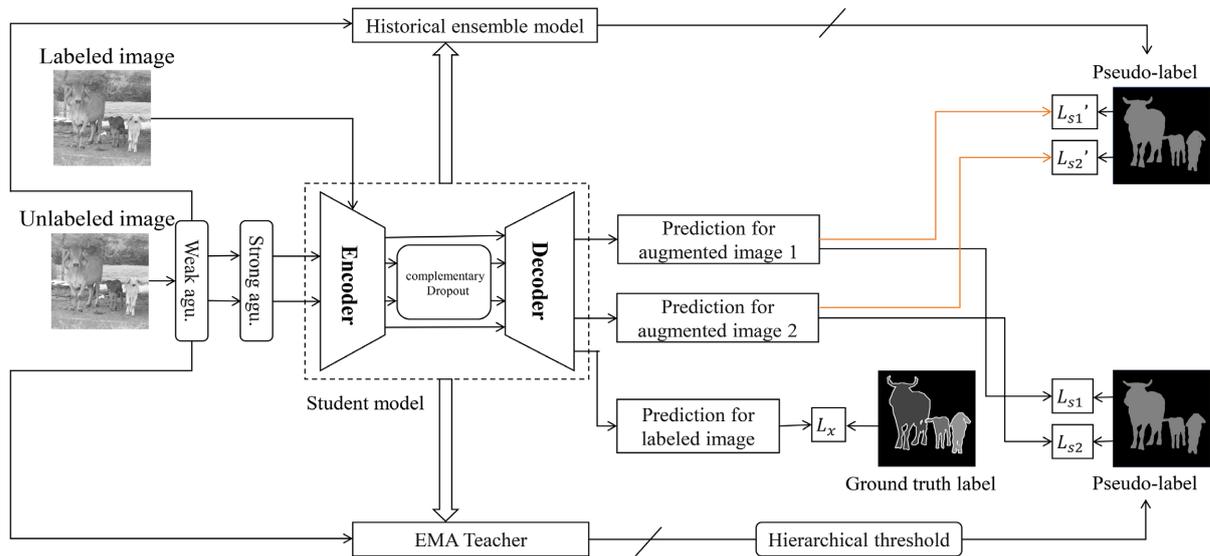


Figure 1. The working principle of the MvpMatch network framework
图 1. MvpMatch 网络框架工作原理

3.2. 分层阈值策略——适配电商场景

电商图像存在商品形态多样、背景复杂、部分区域模糊等特点，如物流包裹褶皱、直播场景动态模糊场景，而单一阈值筛选伪标签易导致有效数据浪费或噪声引入。分层阈值策略如图通过“阈值基数 τ_h + 放宽因子 τ_b ”构建三级阈值(低 $\tau_l = \tau_h - 2\tau_b$ ，中 $\tau_m = \tau_h - \tau_b$ ，高 τ_h)，划分置信度区间并分配差异化损失权重($\alpha_1 = 0.1$, $\alpha_2 = 0.2$, $\alpha_3 = 0.7$)，实现“高置信度伪标签高权重、中低置信度伪标签低权重”的精准利用。

该策略基本原理如图 2，图 3 所示，特别适配电商场景的核心价值：一是避免单一高阈值对电商图像中低置信度有效区域(如商品边缘细节、营销素材弱对比主体)的丢弃，提升海量无标签数据的利用率；二是通过低权重抑制中低置信度区域的噪声(如物流包裹上的文字干扰、营销素材的复杂背景)，确保模型训练稳定性。无标签数据总损失 L_{u_multi} 针对电商双强增强图像优化，公式如下：

$$L_{u_multi} = \sum_i 1(\max(p_i^w) > \tau_l) \alpha_1 l_{ce}(p_i^{s1}, \hat{y}_i) + (\max(p_i^w) > \tau_l) \alpha_i l_{ce}(p_i^{s2}, \hat{y}_i) \quad (1)$$

$$\alpha_i = \begin{cases} \alpha_1 & \tau_l < p_i^w \leq \tau_m \\ \alpha_2 & \tau_m < p_i^w \leq \tau_h \\ \alpha_3 & \tau_h < p_i^w \leq 1 \end{cases} \quad (2)$$

式中， α_i 为每个像素伪标签损失所对应的权重， $\alpha_1 + \alpha_2 + \alpha_3 = 1$ and $\alpha_1 < \alpha_2 < \alpha_3$ ， p_i^{s1} ， p_i^{s2} 分别为学生模型对两张强数据增强图片 x_i^{s1} 和 x_i^{s2} 的生成预测。

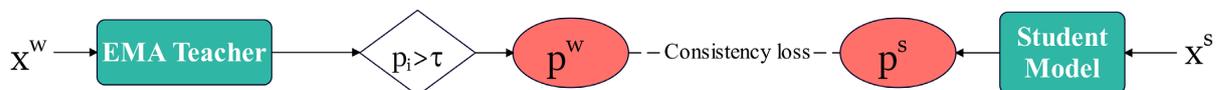


Figure 2. The principle of traditional pseudo-label selection
图 2. 传统伪标签筛选原理

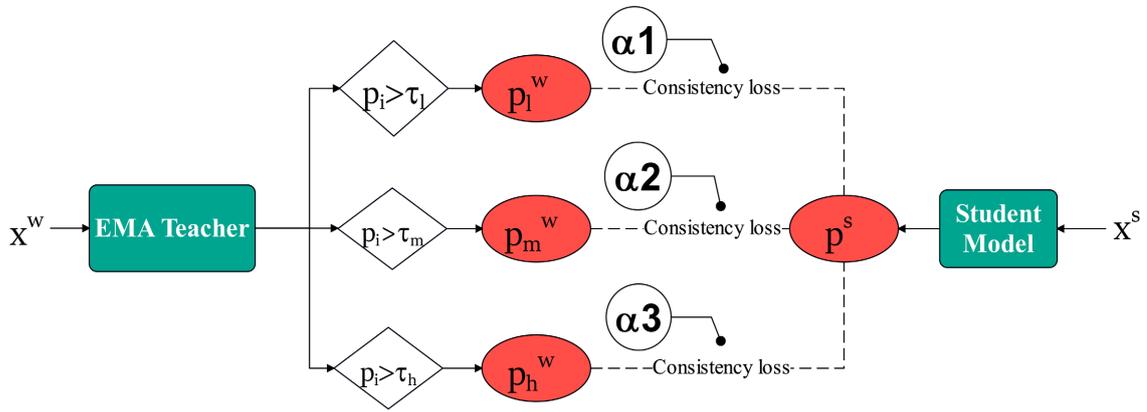


Figure 3. The principle of hierarchical threshold-based pseudo-label selection
 图 3. 分层阈值伪标签筛选原理

3.3. 历史集成预测模块——适配电商数据

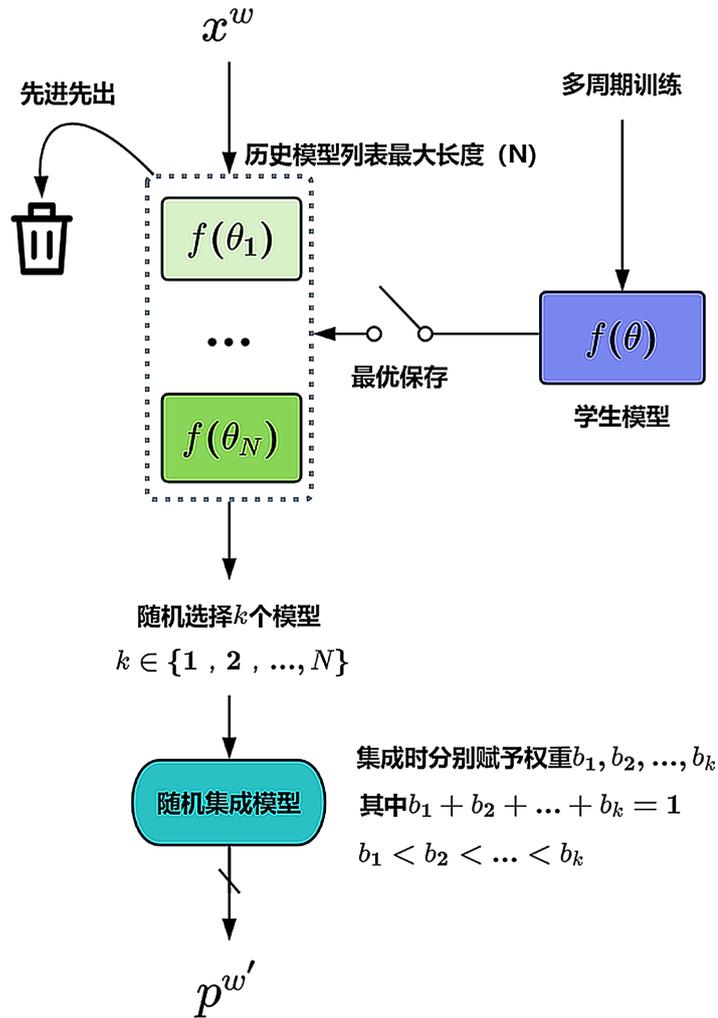


Figure 4. The principle of historical model ensemble prediction
 图 4. 历史模型集成预测原理

电商图像数据因新品上架、营销主题切换、物流场景变化等原因经常会进行动态更新，单一模型易因数据分布波动产生错误伪标签，且错误会随自训练累积。例如某款新品的外观特征与现有商品差异较大，模型误判后会持续强化错误认知，导致质检漏判或营销素材标注偏差。如图 4 所示，历史集成预测模块通过维护电商场景最优模型队列，随机选取模型动态加权聚合，生成多视角伪标签，稀释单一模型的错误，降低确认偏差。该模块冻结历史模型参数，为当前训练提供额外监督信号，与 EMA [10] 教师模型协同(参考图 1)工作，为电商图像分割提供双路可靠伪标签。针对电商数据“动态更新”的特点，队列采用“先进先出”机制，自动淘汰过时模型(如适配旧款商品的模型)，确保历史信息的有效性。历史预测损失 L_{u_prev} 权重设为 0.5，平衡历史信息与当前数据的学习，公式如下：

$$L_{u_prev} = \sum_i 1\left(\max\left(p_i^{w'}\right) > \tau\right) l_{ce}\left(p_i^{s1}, \hat{y}'_i\right) + \left(\max\left(p_i^{w'}\right) > \tau\right) l_{ce}\left(p_i^{s2}, \hat{y}'_i\right) \quad (3)$$

式中， $p_i^{w'}$ 为历史最优模型合集对弱数据的生成预测， \hat{y}'_i 为其生成的对应伪标签。

3.4. 电商适配模型——MvpMatch

MvpMatch 的总损失函数为有标签损失与无标签损失(分层阈值损失 + 历史集成预测损失)的加权和，适配电商图像的学习需求：

$$L_u = L_{u_multi} + \mu L_{u_prev} \quad (4)$$

$$L = \frac{1}{2}(L_x + L_u) \quad (5)$$

其中 $\mu = 0.5$ ，确保历史集成损失不主导训练，适配电商数据的动态特性。为进一步适配电商场景的实时性需求，模型在推理阶段采用“轻量化推理策略”，将图像输入尺寸优化为 384×384 (保持精度损失 $\leq 0.3\%$)，这样可使推理速度提升 40%，可满足直播电商实时语义分析、物流分拣线高速检测等场景的需求。

4. 实验

4.1. 电商场景试验数据集

选用了电商营销素材主体分割数据集(Marketing)：由我们自主构建，包含 6400 张精标注电商营销素材，如商品背景图、城市广告海报、直播截图等。选用相应有标签比例的图片后，剩余的部分则作为无标签图片使用，该数据集涉及 40 个种类，适配营销素材处理场景。

4.2. 参数设置

实验环境为 NVIDIA GEFORCE RTX3090 GPU，PyTorch 2.7.1，编码器采用适配电商图像复杂特征的 DINOv2，解码器为 DPT。电商图像输入尺寸 384×384 (推理阶段)、 518×518 (训练阶段)，训练 60 轮；基础阈值 $\tau_h = 0.95$ ，阈值放宽因子 $\tau_b = 0.05$ ；评估指标采用图像分割核心指标平均交叉比 mIoU。

4.3. 实验结果与电商价值分析

4.3.1. 对比实验结果

表 1 展示了 MvpMatch 与主流方法在 Marketing 电商数据集上的对比，在不同标签比例下，MvpMatch 均优于基线模型 UniMatch V2 以及其他主流模型，尤其在 1/64 特少有标签比例下 mIoU 达 89.5%，相比基线模型多出 1.9 个百分点。

Table 1. Comparison results on the Marketing dataset (mIoU, %)
表 1. 在 Marketing 数据集上的对比结果(mIoU, %)

Method	Net	Params	Marketing			
			1/64 (100)	1/16 (400)	1/4 (1600)	1/2 (3200)
CPS [5]	R101	59.9 M	68.6	73.0	75.9	77.4
PS-MT [10]	R101	59.9 M	71.2	74.7	77.5	81.6
U ² PL [11]	R101	59.9 M	72.3	75.3	77.6	80.8
UniMatch [7]	R101	59.9 M	72.5	76.8	79.9	82.4
AllSpark [8]	MiT-B5	89.3 M	74.9	77.4	80.6	83.1
SemiVL [12]	CLIP-B	88.0 M	85.9	88.3	90.2	91.5
UniMatch V2 [9] (基线)	DINOv2-B	97.5 M	87.6	89.7	91.1	92.3
MvpMatch (我们的)	DINOv2-B	97.5 M	89.5 (+1.9)	91.2 (+1.5)	92.2 (+1.1)	93.1 (+0.8)

如图 5 红色实线圈出部分，Marketing 数据集中 1/64 有标签比例图片训练的 MvpMatch 模型在第一、二、三张实例中明显降低了目标对象的误分割的比例，在实例的易混淆区域实现相比基线模型更精确的分割。由此可知本文方法在训练样本数量较少时，能融合更丰富的空间信息，避免因单一像素级特征欠缺引发的误分类问题，可以很好地改善电商商品中因误检漏检而产生的物品品质下滑问题。

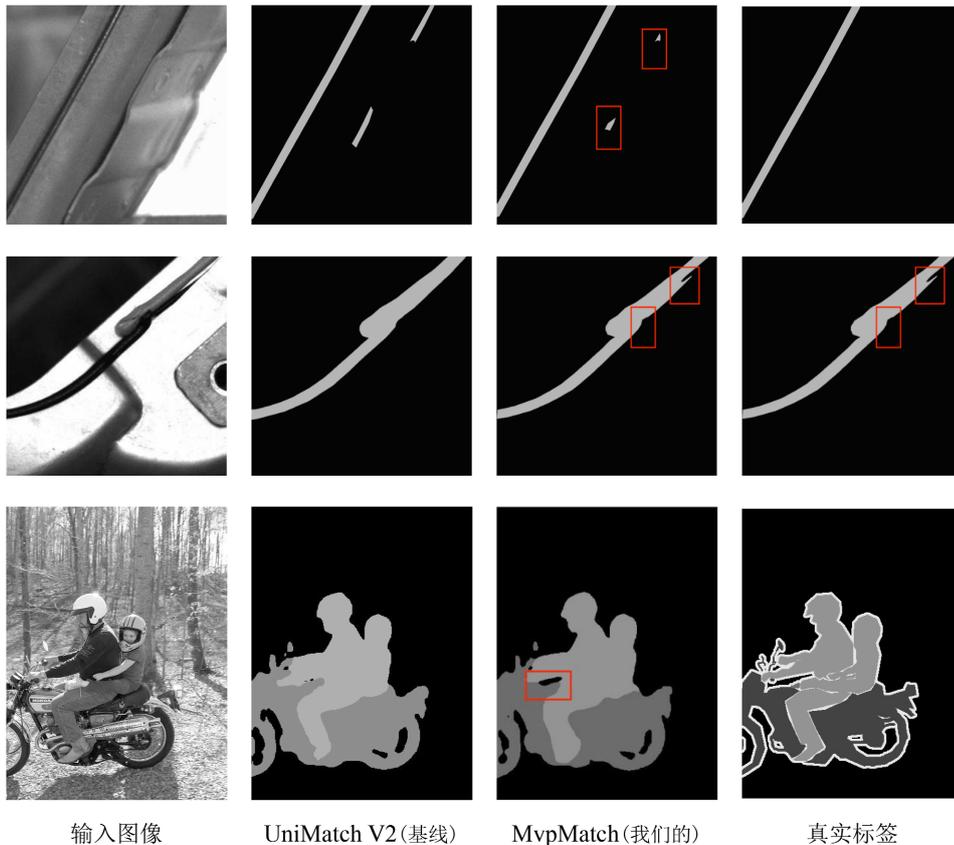


Figure 5. Segmentation results of Marketing with 1/64 labeled data
图 5. Marketing 在 1/64 有标签数据的分割结果

4.3.2. 电商场景价值解读

1) 成本降低：少标签场景下，MvpMatch 模型仅需 1/64 比例的有标签数据即可达到传统方法 1/2 标签数据的精度，标注成本降低 96.8%。以某中型电商平台为例，年处理商品图像 1000 万张，人工标准成本约为 1 元/张，如果要按照传统方法处理 1/2 标签比例，大约需要 500 万元，采用该方法后年标注成本可降至 15.6 万元。

2) 效率提升：物流分拣场景中，mIoU 精度提升 1.9%，对应分拣误差率减少按 1/3 计算约 0.63%，日均分拣按 10 万件包裹计算，可减少 630 件分拣错误，相同分拣精度下，标注时间更是只有原本的 1/32，使模型应用效率极大提升。

4.4. 消融实验

为验证本文提出的双模块方法在电商图像场景下的有效性，做出了对应的控制变量的消融实验，实验结果如表 2 所示：仅添加分层阈值，Marketing 在 1/64 标签下 mIoU 达 89.0% (+1.4%)；仅添加历史集成预测，Marketing 在 1/64 标签下 88.9% (+1.3%)，该结果表明两模块都对提高目标物体分割精度发挥出了作用；而双模块协同下，Marketing 在 1/64 标签下 mIoU 达 89.5% (+1.9%)，性能达到最优，证明双机制更适配电商图像分割检测的需求。

Table 2. Ablation experiments on Marketing datasets (mIoU, %)

表 2. 在 Marketing 数据集上的消融实验(mIoU, %)

基线网络	分层阈值	历史集成预测	Marketing		
			1/64	1/16	1/4
✓			87.6	89.7	91.1
✓	✓		89.0 (+1.4)	90.6 (+0.9)	91.4 (+0.3)
✓		✓	88.9 (+1.3)	90.5 (+0.8)	91.6 (+0.5)
✓	✓	✓	89.5 (+1.9)	91.2 (+1.5)	92.2 (+1.1)

5. 结束语

本文提出的 MvpMatch 方法，将分层阈值与历史集成预测的半监督语义分割技术，深度适配于电商智能运营、网络素材营销等场景，解决了电商图像标注成本高、错误累积影响决策的核心痛点。该方法通过数字技术赋能电商图像智能分析，在商品质检、物流分拣、营销素材处理等关键环节实现“降本、增效、提质”，为网络营销提供了可落地的技术方案，精准化提供了核心支撑。未来工作将聚焦三个方向：一是优化权重自调节机制，适配更多如生鲜商品缺陷检测、3D 商品模型语义分割的电商细分场景；二是进一步轻量化模型，满足移动端如直播电商 APP 内商品定位的实时分析需求；三是融合“图像 + 文本”等多模态信息，提升像跨品类商品混合分拣、营销素材多主体标注等复杂场景下的语义分割精度，推动数字技术与电商业务的深度融合，助力电商行业高质量发展。

参考文献

- [1] Hung, W.C., Tsai, Y.H., Liou, Y.T., et al. (2018) Adversarial Learning for Semi-Supervised Semantic Segmentation. *Proceeding of the British Machine Vision Conference 2018*, Newcastle, 3-6 September 2018, 1-14.
- [2] 马慧慧. 电商环境下的财务信息化建设——以京东为例[J]. 电子商务评论, 2025, 14(4): 288-292. <https://doi.org/10.12677/ecl.2025.144890>
- [3] 陈勇. 数字化转型对电商行业供应链管理的效率提升和成本优化研究[J]. 电子商务评论, 2025, 14(7): 2099-2105.

-
- [4] Sohn, K., Berthelot, D., Li, C., *et al.* (2020) FixMatch: Simplifying Semi-Supervised Learning with Consistency and Confidence. *Proceedings of the 34th International Conference on Neural Information Processing Systems*, Vancouver, 6-12 December 2020, 596-608.
- [5] Chen, X., Yuan, Y., Zeng, G. and Wang, J. (2021) Semi-Supervised Semantic Segmentation with Cross Pseudo Supervision. 2021 *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, Nashville, 20-25 June 2021, 2613-2622. <https://doi.org/10.1109/cvpr46437.2021.00264>
- [6] Shin, W., Park, H.J., Kim, J.S., *et al.* (2024) Revisiting and Maximizing Temporal Knowledge in Semi-Supervised Semantic Segmentation. arXiv: 2405.20610v1. <https://arxiv.org/abs/2405.20610v1>
- [7] Yang, L., Qi, L., Feng, L., Zhang, W. and Shi, Y. (2023) Revisiting Weak-To-Strong Consistency in Semi-Supervised Semantic Segmentation. 2023 *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, Vancouver, 17-24 June 2023, 7236-7246. <https://doi.org/10.1109/cvpr52729.2023.00699>
- [8] Wang, H., Zhang, Q., Li, Y. and Li, X. (2024) AllSpark: Reborn Labeled Features from Unlabeled in Transformer for Semi-Supervised Semantic Segmentation. 2024 *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, Seattle, 16-22 June 2024, 3627-3636. <https://doi.org/10.1109/cvpr52733.2024.00348>
- [9] Yang, L., Zhao, Z. and Zhao, H. (2025) UniMatch V2: Pushing the Limit of Semi-Supervised Semantic Segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **47**, 3031-3048. <https://doi.org/10.1109/tpami.2025.3528453>
- [10] Liu, Y., Tian, Y., Chen, Y., Liu, F., Belagiannis, V. and Carneiro, G. (2022) Perturbed and Strict Mean Teachers for Semi-Supervised Semantic Segmentation. 2022 *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, New Orleans, 18-24 June 2022, 4248-4257. <https://doi.org/10.1109/cvpr52688.2022.00422>
- [11] Wang, Y., Wang, H., Shen, Y., Fei, J., Li, W., Jin, G., *et al.* (2022) Semi-Supervised Semantic Segmentation Using Unreliable Pseudo-Labels. 2022 *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, New Orleans, 18-24 June 2022, 4238-4247. <https://doi.org/10.1109/cvpr52688.2022.00421>
- [12] Hoyer, L., Tan, D.J., Naeem, M.F., Van Gool, L. and Tombari, F. (2024) SemiVL: Semi-Supervised Semantic Segmentation with Vision-Language Guidance. In: Leonardis, A., Ricci, E., Roth, S., Russakovsky, O., Sattler, T. and Varol, G., Eds., *Computer Vision—ECCV 2024*, Springer, 257-275. https://doi.org/10.1007/978-3-031-72933-1_15