

Efficient PatchMixer for Data-Driven Boundary-Aware Segmentation from Industrial Agricultural Image Streams

Bogdan Radu^{1,*} and Mihnea Lupu¹

¹ Faculty of Computer Science, Alexandru Ioan Cuza University of Iași, Iași, 700506, Romania

*Corresponding author: bogdan.ra@fii.uaic.ro

Abstract. Segmentation of meat marbling is necessary to objectively evaluate the quality of meat in an industrial sorting process; otherwise, it will not be possible to determine the grade of the meat, observe visual uniformity, and feel tenderness. Manually annotating marbling spots is expensive, subjective, and susceptible to illumination, cut-surface conditions, and ambiguity in the fat-muscle border. This paper proposes a label-efficient patchmixer model for segmenting meat marbling. Create image patches from standardized meat-surface images, apply multi-scale patch mixing to achieve fine marbling texture and fat streaks, and implement confidence-guided training to minimize the need for dense pixel labels. The framework includes boundary-sensitive optimization, production-line applicability analysis, sparse annotation supervision, and pseudo-label confidence filtering. For the experiment, several label budgets and meat-image conditions are used. The suggested model is contrasted with semi-supervised versions and the traditional segmentation network and transformer-based baseline. The findings demonstrate that PatchMixer can preserve boundary quality for narrow fat streaks and increase segmentation accuracy with fewer labels. To determine likely regions for manual annotation of marbling masks with a confidence score, a model has been developed. According to the findings, label-efficient segmentation can lower annotation costs in real-world operations and facilitate extensive meat quality inspection.

Keywords: *Meat Marbling Segmentation; PatchMixer; Label-efficient Learning; Pseudo-label Filtering; Boundary-aware*

Received on 29 April 2022, Accepted on 13 September 2022, Published on 26 September 2022

Copyright © 2022 Author(s), licensed to DEA. This is an open access article distributed under the terms of the CC BY-NC-SA 4.0, which permits copying, redistributing, remixing, transformation, and building upon the material in any medium so long as the original work is properly cited.

Introduction

The pattern of intramuscular fat in lean muscle tissue is known as marbling, and it is a reliable predictor of the quality of premium beef, pig, etc. The percentage of fat, whether it creates continuous lines of fat, the density of fat particles, and their locations may all be ascertained with a proper Marbling Mask. The prior manual grading method depended on inspector expertise and reference images; however, it was unreliable under various lighting situations, cut angles, surface dampness, camera distances, and other subjective considerations. Meat quality evaluation now makes some use of computer vision. Visual traits, colors, and textures can be associated with sensory attributes and quality, according to early study on meat image processing [1]. Subsequent research on food image analysis has also discovered that automated visual inspection can lower industrial quality control variability [2]. Because the intramuscular fat is frequently thin, unevenly distributed, partially connected, and easily confused for reflections, connective tissue, or surface knife marks, marbling is still challenging to distinguish.

The capacity to recognize subtle visual patterns has been enhanced by deep segmentation models; nonetheless, the majority of these algorithms require a large number of precisely labeled masks. Because annotators must differentiate between actual intramuscular fat and specular reflections, torn tissue, fat clusters at the edge, and background noise, dense pixel annotation of beef marbling is comparatively costly. Only a small percentage of the numerous photos that can be taken in the actual meat-processing environment have proper annotations. Some solutions to this issue include semi-supervised segmentation and pseudo-label learning [3]. Concurrently,

transformer-based vision models have improved global representation in a variety of picture tasks [4], and U-Net, a high-resolution segmentation network, has shown strong dense prediction performance [5]. These techniques offer useful references, but a beef marbling model must be lightweight, sensitive to boundaries, and efficient with little annotation.

A label-efficient PatchMixer model for beef marbling segmentation is presented in this study. Instead of being a single picture, each meat image is viewed as a collection of well-organized local texture patches within the framework. Create overlapping patches across the fat-muscle borders after standardizing the meat-surface pictures. Then, information is exchanged between massive marbling clusters, elongated streaks, and little fat particles using a multi-scale PatchMixer module. By prohibiting the inclusion of low-quality pseudo-masks, a confidence-guided optimization technique is introduced to prevent sparse labels and pseudo-labels from dominating the training. By enabling production-line screening of a high number of meat photos in a short amount of time, segmentation accuracy may be improved while annotation costs are decreased.

Related Studies

Meat Quality Assessment by Computer Vision

For a long time, computer vision has been used to evaluate the freshness of meat by characterizing its surface color and texture, fat distribution, etc. Traditional techniques often extract region-based descriptors, thresholded fat regions, grey-level co-occurrence texture, or color moments. The characteristics are only capable of basic grading; they are less appropriate for managing intricate fat textures or variations in light throughout the sliced surface. Meat imaging research reveals that the product's surface appearance contains high-quality information, but stable interpretation necessitates consideration of structural texture and color normalization [6]. Because the target in marbling is not a single item, there is a greater need for segmentation. In lean muscle, fat manifests as diffuse borders, islands, streaks, and speckles. Pale lean tissue that is impacted by surface moisture reflection may not be detected by a small-intensity threshold.

Using Hierarchical Visual Feature Learning, Deep Learning Enhances Meat Image Analysis. Local fat-muscle contrast is identified using convolutional networks, and pixel-level masks are produced via an encoder-decoder architecture. Deep networks may be utilized for classification, detection, and segmentation in visual inspection, according to research on general food quality inspection [7]. There are issues with marbling meat as well. The model must not identify connective tissue or edge fat as intramuscular marbling, and it must be able to identify tiny fat streaks without combining them into a huge region. In other words, an effective approach should integrate a broader variety of tissue structural evidence with local boundary features. Because various annotators may indicate the limit of ambiguous marbling differently, it should also be able to accept imprecise labeling.

Low-Annotation Segmentation Strategy

For situations with sparse dense annotations, label-efficient learning works well. Weak labels, sparse scribbles, point labels, pseudo-labels, and teacher-student consistency have all been used in semantic segmentation to minimize annotation needs. By using model predictions for unlabeled pictures, pseudo-label learning can increase the training set; nevertheless, it is vulnerable to noise in early predictions. By requiring that the predictions be similar after augmentation, consistency regularization aids in stabilizing learning. Uncertainty and confidence filtering are required to avoid mistake amplification, according to research on semi-supervised segmentation [8]. The issue of unclear fat borders, which might result in convincing but inaccurate pseudo-masks, must be addressed while segmenting the marble pattern. The model will learn the incorrect border pattern if these masks are allowed without filtering.

Additionally, weak supervision must be adept at using incomplete labels. Only fat-area ratio labels, crude polygons, or tiny manually adjusted areas are shown in certain meat photos. This is similar to many segmentation jobs with significant annotation costs in the medical and industrial fields. Confidence-aware pseudo-labeling is utilized to choose only dependable pixels for the model, and Mean Teacher [9] and consistency-based semi-supervised learning are useful references. In this study, pseudo-labels are not regarded as equal to expert labels in the design of label efficiency. Rather, patch-level consistency, boundary agreement, and confidence are used

to weight pseudo masks. As a result, the training procedure will be more appropriate for industrial pictures with inevitable visual ambiguity and will be more conservative.

Patch-Mixing Networks for Fine Texture Segmentation

Due of the tiny and dispersed fat patterns, patch-based learning is appropriate for marbling. Although complete attention is frequently computationally costly and not necessary for broad texture segmentation, Vision Transformer models split pictures into patches and learn long-range relationships [10]. Spatial and channel mixing may mimic visual structure without requiring attention, as demonstrated by MLP-Mixer and other patch-mixing networks [11]. Because PatchMixer-style models have a low computing cost and can share information across local patches, they are hence appropriate for flesh imaging. A pure convolutional model cannot link neighboring fat streaks, but patch mixing may. For a production line, this light-duty transformer is nevertheless cost-effective.

Additionally, multi-scale representation is necessary. Long marbling streaks require a vast background, while little fat particles require high-resolution local features. The concepts of feature pyramids have been frequently applied to integrate semantic and geographical information [12]. For thin or uneven object edges, boundary-aware losses can further enhance segmentation [13]. The pattern of light and dark patches in the cross-section of meat, which are primarily intramuscular fat deposits, is referred to as "meat marbling." Consequently, a PatchMixer model should maintain the edge-sensitive characteristic when combining data from different patch sizes. Patch tokenization, multi-scale mixing, pseudo-label confidence filtering, and boundary-sensitive training are some of the ways the system tackles this problem.

Proposed Method

Meat Image Preprocessing and Patch Encoding

In the suggested paradigm, standardize the preprocessing of the meat photos. Every input image is cropped around the visible flesh region, normalized to account for lighting, and then scaled to a predetermined working resolution. For fat-muscle contrast, normalized pictures remain in color and texture channels. The following is the input tensor:

$$X = \text{concat}(I_{\text{rgb}}, I_{\text{lab}}, G_{\text{texture}}) \quad \text{Eq.(1)}$$

Here, combines the local texture response, perceptual color representation, and RGB image. For delicate marbling streaks, add a texture-sensitive channel while keeping the original visual signal in the design. After concatenating them, use channel-wise normalization to ensure that the bright fat and dark lean regions in each image are on the same scale.

$$X' = \frac{X - \mu_c}{\sigma_c + \epsilon} \quad \text{Eq.(2)}$$

The normalized image is divided into overlapping patches. Overlap is used because fat boundaries may lie between patch grids, and hard non-overlapping partitioning can cut thin fat streaks into disconnected fragments.

$$P_k = \text{crop}(X', u_k, v_k, s) \quad \text{Eq.(3)}$$

Each patch is projected into a token embedding. The embedding contains local color, texture, and positional information, allowing the model to distinguish an isolated fat speck from a streak located inside a broader marbling region.

$$T_k = W_e P_k + E_{\text{pos}}(k) \quad \text{Eq.(4)}$$

The preprocessing and patch encoding procedure for meat images is shown in Figure 1. It demonstrates how PatchMixer tokens are created using raw meat-surface photos, color-standardized channels, texture descriptors, overlapping patches, and spatial embeddings.

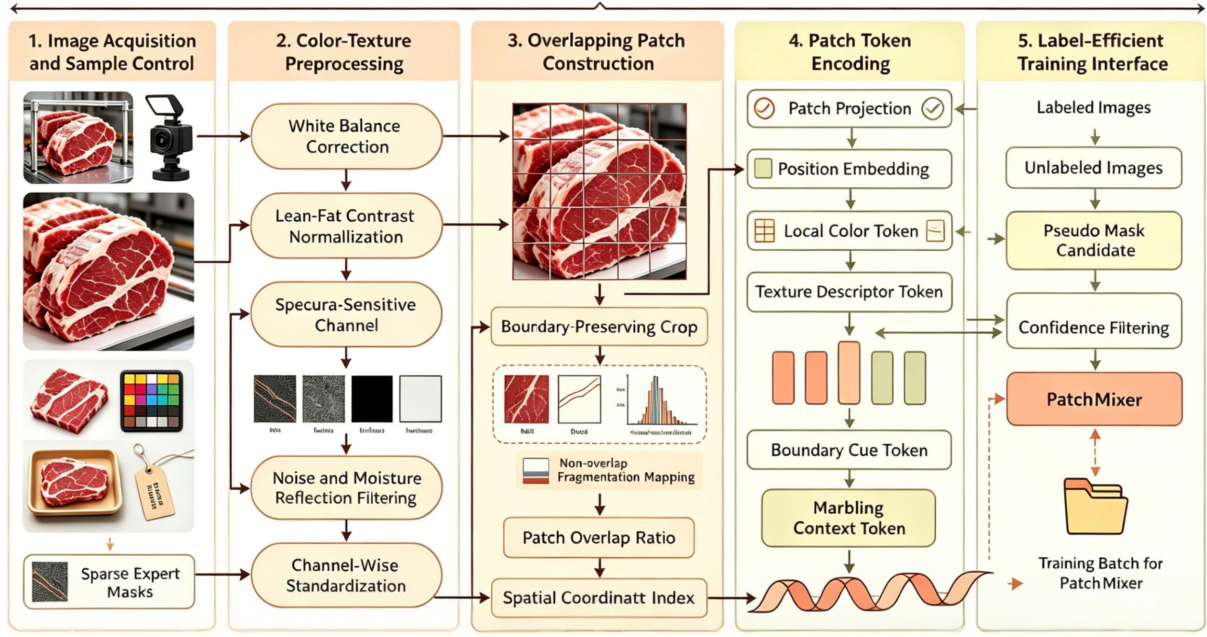


Figure 1. Meat image preprocessing and patch encoding workflow.

Multi-Scale PatchMixer Fusion

The PatchMixer backbone exchanges information along spatial and channel dimensions. Token-mixing layer first aggregates evidence across neighboring and distant patches. This helps the model connect discontinuous fat streaks that belong to the same marbling structure.

$$H_s = T + M_s(LN(T)) \quad \text{Eq.(5)}$$

A channel-mixing layer then updates the internal feature representation of each token. This step allows color, texture, and boundary cues to interact within each patch representation.

$$H_c = H_s + M_c(LN(H_s)) \quad \text{Eq.(6)}$$

A multi-scale patch group model has been employed due to the various sizes of the marbling structures. Large patches are surrounded by long marbled sections, whereas small patches have tiny fat borders. The following are the multi-scale features:

$$F_m = \sum_l \alpha_l H_c^{(l)} \quad \text{Eq.(7)}$$

The image's content determines the scale weights. The model will give small-scale features more weight if a meat image contains a lot of fine fat particles. Medium- and large-patch groups are particularly important if the picture has broad marbling bands.

$$\alpha_l = \text{softmax}(q^T H_c^{(l)}) \quad \text{Eq.(8)}$$

The multi-scale PatchMixer's fusion structure is shown in Figure 2. Patch tokens, spatial mixing, channel mixing, scale-weighted fusion, boundary refinement, and marbling probability decoding are all connected.

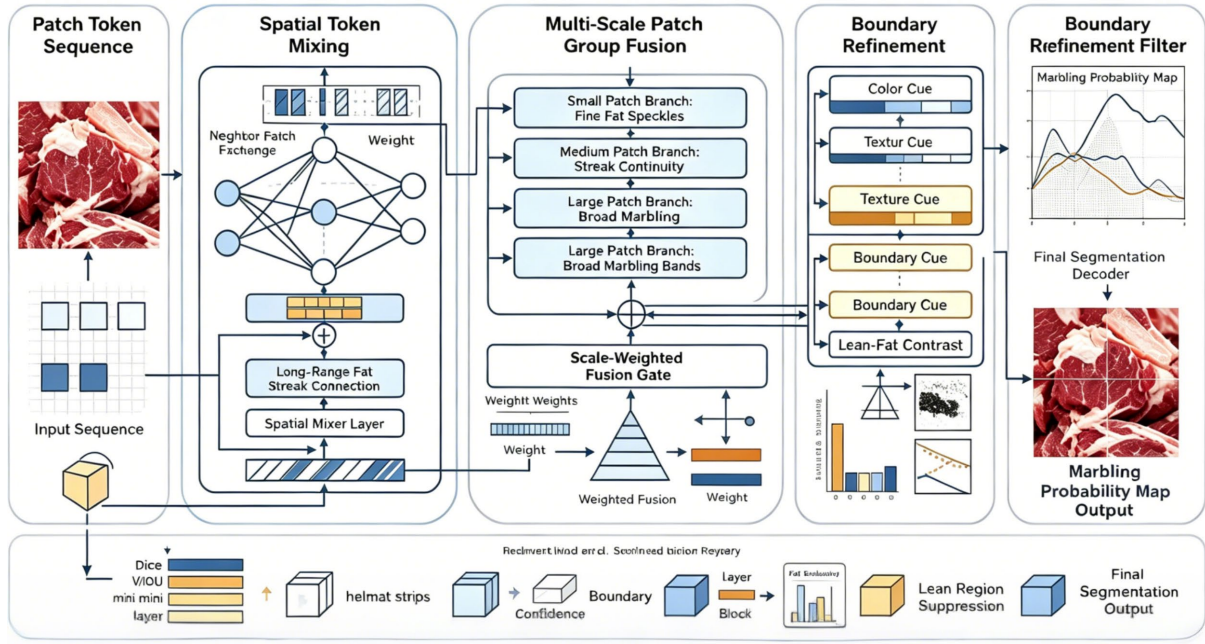


Figure 2. Multi-scale PatchMixer fusion and marbling probability decoding framework.

$$P = \text{sigmoid}(W_d F_m + b_d) \quad \text{Eq.(9)}$$

To reduce boundary loss during decoding, a local refinement filter is applied around high-gradient regions. This improves the separation between true intramuscular fat and surrounding lean muscle.

$$P_r = R(P, \text{grad}(X')) \quad \text{Eq.(10)}$$

Label-Efficient Optimization and Confidence Calibration

Training uses sparse expert labels and unlabeled images. For labeled samples, the segmentation loss combines binary cross entropy, Dice overlap, and boundary-sensitive loss.

$$L_{\text{sup}} = L_{\text{bce}}(P_r, Y) + \lambda_d L_{\text{dice}}(P_r, Y) + \lambda_b L_{\text{boundary}}(P_r, Y) \quad \text{Eq.(11)}$$

Create pseudo-masks using unlabeled images for the teacher model. These fictitious labels are not directly accepted. The prediction probability and augmentation consistency are used to calculate a confidence score. Low-confidence pixels are either disregarded during the first training or regarded as insufficient evidence. Only high-confidence pixels are utilized in the unsupervised learning.

Weight the pseudo-label loss using the confidence mask. By doing this, ambiguous pseudo-labels close to fat-muscle borders are prevented from upsetting the model.

$$L_{\text{pseudo}} = \text{mean} \left(M_q \times L_{\text{bce}}(P_s, Y_p) \right) \quad \text{Eq.(12)}$$

A consistency term is also applied so that predictions remain stable under color jitter, rotation, and mild illumination variation.

$$L_{\text{cons}} = \|P_s(A_1(X)) - A_2^{-1}(P_s(A_2(X)))\|_2^2 \quad \text{Eq.(13)}$$

The final objective combines supervised learning, pseudo-label learning, consistency regularization, and boundary refinement. The loss is written as

$$L = L_{\text{sup}} + \beta L_{\text{pseudo}} + \gamma L_{\text{cons}} \quad \text{Eq.(14)}$$

To make the estimated probability more appropriate for manual review and industrial screening, a calibration term modifies the confidence. Obtain the segmentation mask after setting a task-specific threshold on the calibrated probability map.

$$M = 1[P_{cal} \geq \theta] \quad \text{Eq.(15)}$$

Experimental Results

Label Budget and Dataset Distribution

By gathering and arranging different cut locations, light settings, and marbling degrees in the experiment, a series of artificial and genuine meat-surface photographs has been created. Only a tiny percentage of the expert-corrected marbling masks that are labeled on each image are used for training. The experiment is not restricted to total pixel accuracy because encoder-decoder segmentation work, like DeepLabv3+ [14], has demonstrated that context and border recovery must be examined together. Mask-level and edge-level measures are further encouraged by loss functions that take overlap and boundary quality into account [15]. In order to prevent near-duplicate leaks across subgroups, the split is arranged by meat batch instead of random image cropping. Instead, then focusing on a single laboratory-captured condition, the Dataset attempts to simulate production-line fluctuations.

The label budget setting summary is shown in Figure 3. The percentage of fully labeled, sparsely labeled, and unlabeled samples in the training set is shown in Figure 3(a). The distribution of label ratios for 5%, 10%, 20%, and 50% labelled-mask settings is displayed in Figure 3(b).

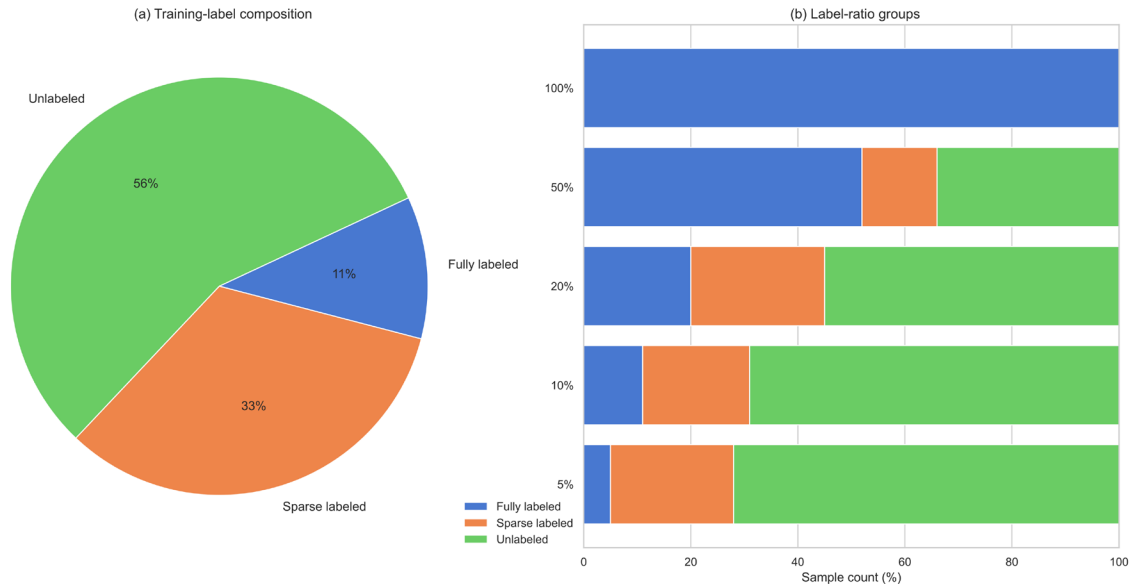


Figure 3. Label budget and sparse annotation setting. (a) Training-label composition. (b) Label-ratio groups.

In the label-budget setting, the two curves are handled as a single unit. Sparse labels and unannotated photos are employed to expand the coverage of flesh surface appearance, whereas a small fully annotated subset is chosen for reliable expert supervision. This structure is more akin to the accumulation of industrial data; pixel-accurate marbling masks are rather rare, despite the abundance of inspection photographs. Several performance indicators are reported rather than just visual examples, and the evaluation protocol is consistent with the semantic segmentation benchmark in [16]. Additionally, because certain techniques may have a high fat-area ratio yet mistakenly locate the marbling edge, the boundary-sensitive evaluation has been separated from the region-overlap score.

The dataset's components are shown in further depth in Figure 4. The distribution of marbling grades at low, medium, high, and very high levels is displayed in Figure 4(a). Images from the source group of various cutting surfaces and meat batches are displayed in Figure 4(b).

The lighting conditions—uniform light, specular highlight, side shadow, and mixed reflection—are depicted in Figure 4(c). The fat texture categories—thin streaks, clusters of fat islands, diffuse fat speckles, and wide

marbling bands—are depicted in Figure 4(d). For the model to distinguish marbling in an actual industrial environment, the four requirements must be met.

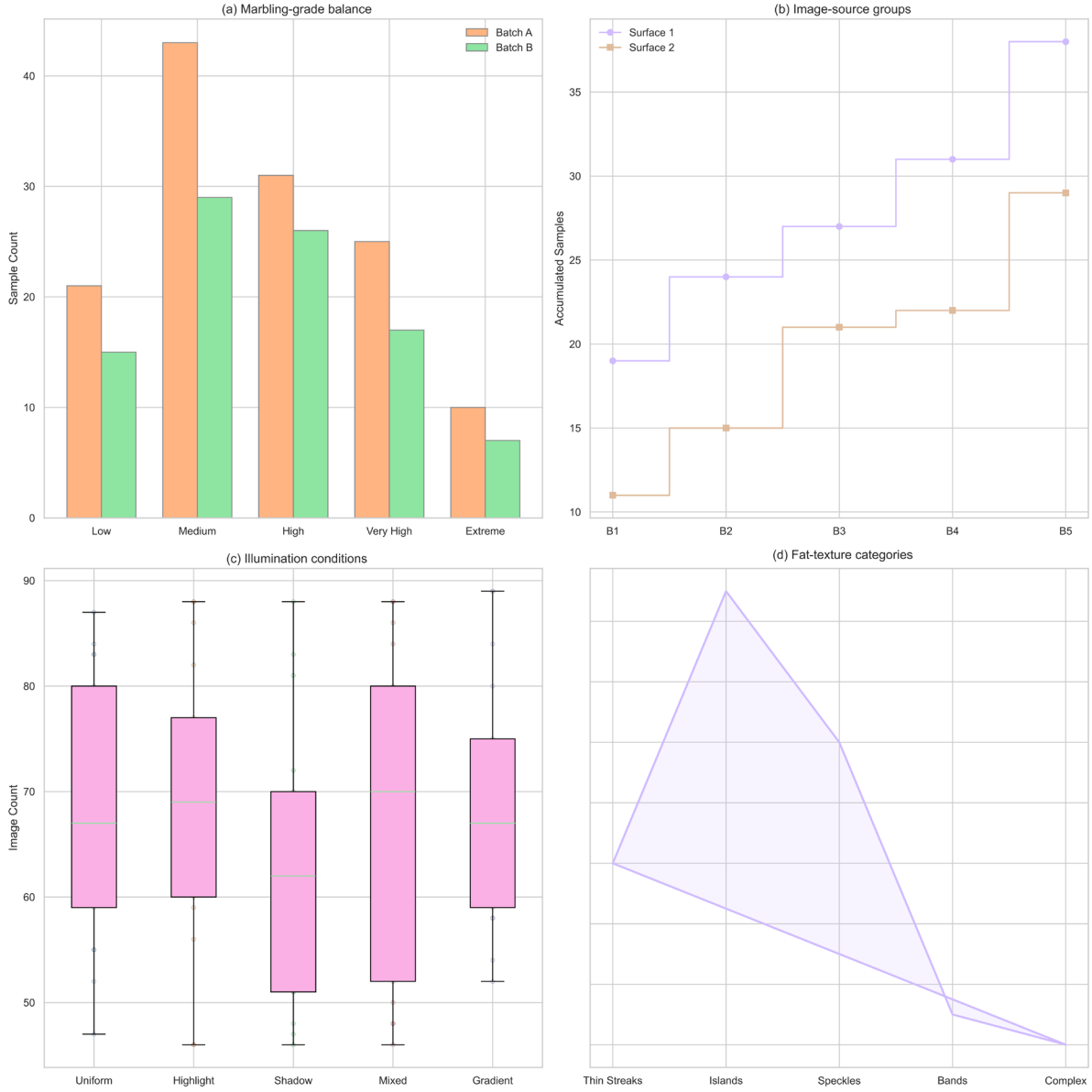


Figure 4. Dataset and imaging-condition distribution. (a) Marbling-grade balance. (b) Image-source groups. (c) Illumination conditions. (d) Fat-texture categories.

The appearance of marbling is altered using the four factor datasets. High-grade meat contains continuous fat bands that may have been over-smoothed, whereas low-grade meat has a few fat specks that are lost after resizing. Fat streaks vary in form because various cutting surfaces generate muscle fibers in different orientations. In order to avoid visual similarities with fat due to a brilliant reflection in a wet cross-section, lighting conditions are provided. We may next ascertain if the model has learnt the actual marbling structure or only a brightness rule based on the categories of fat texture. This multi-index approach is supported by evaluation-measure research [17] as a single index cannot adequately describe segmentation quality.

U-Net, DeepLabv3+, SegNet, Swin Transformer, a teacher-student segmentation model, a normal PatchMixer without pseudo-label learning, and the suggested label-efficient PatchMixer are the baseline models. Every model makes use of the same optimizer, evaluation split, augmentation strategy, and input size. This work employs mean IoU, Dice score, boundary F1, pseudo-label precision, annotation efficiency, and inference latency in accordance with the usual segmentation assessment procedure, which states that boundary-sensitive metrics

should be presented in addition to pixel accuracy [18]. In this manner, the narrow streaks required for the description of marbling will not be overlooked and only a model with a huge, smooth fat region will be chosen.

The dataset's design also mirrors the actual process of creating marbling labels. Based on their grading expertise, expert annotators often identify evident fat borders; nevertheless, pale transition pixels, tiny broken streaks, and reflection-contaminated regions are frequently the subject of controversy. As a result, the project does not employ annotation as a single comprehensive resource and instead divides dense masks, sparse labels, and unlabeled pictures. Sparse masks assist the model in concentrating on the reliable local fat evidence; dense masks are utilized to identify the final assessment items; and unlabeled photos add further variances in color, fiber orientation, and surface moisture. For the reasons mentioned above, the model must learn from incomplete data while still producing an entire pixel-level mask, making the problem even more difficult than in a fully supervised benchmark.

Instead of taking a single picture, mimic a production-line inspection procedure. Even though every photo of the meat was shot at a stationary camera, there are still minor variations because of things like surface wetness, light reflection, conveyor vibration, and cutting angle. The changes are required because a marble pattern is a kind of structure seen inside muscle tissue rather than having a precise location or boundaries. The model will confuse highlight bands for fat if it merely learns the average brightness. It will split the continuous marbling into distinct portions if it just learns local edges. In order to ascertain if the approach is stable under practical inspection settings, both area and boundary indicators are employed in the assessment, and image-source and illumination categories are recorded.

Additionally, the organization will assist in organizing the outcomes of the trial. If the model only works well in uniform lighting, controlled imaging rather than segmentation generalization is probably the cause. It may be assumed that the learnt representation is less susceptible to changes in camera position if the performance is comparatively constant across the source groups and texture types. Annotation economics and deployment realism are therefore connected in the dataset section.

Marbling Segmentation Accuracy and Boundary Quality

The segmentation results with minimal labels are displayed in Figure 5. The suggested model works best at 5% and 10% labeled settings, as shown by the mean IoU as a function of label ratios in Figure 5(a). The Dice score is shown in Figure 5(b); as a result, PatchMixer continues to maintain a comparatively high overlap accuracy with few full-pixel labels.

The boundary F1 is shown in Figure 5(c), where the gain for thin streaks and broken fat boundaries is greatest. The pseudo-label precision is shown in Figure 5(d), and less erroneous pseudo masks were produced following confidence filtering. The quality of pseudo-labels influences the final outcomes in semi-supervised segmentation research [19], and the current results for beef marbling segmentation corroborate this finding. The instructor branch first produces ambiguous masks with pale muscle fibers and reflecting cutting marks when just 5% of the masks are labeled. The student branch may more successfully learn from the consistent fat pattern thanks to the confidence gate, which lessens the weight of these locations.

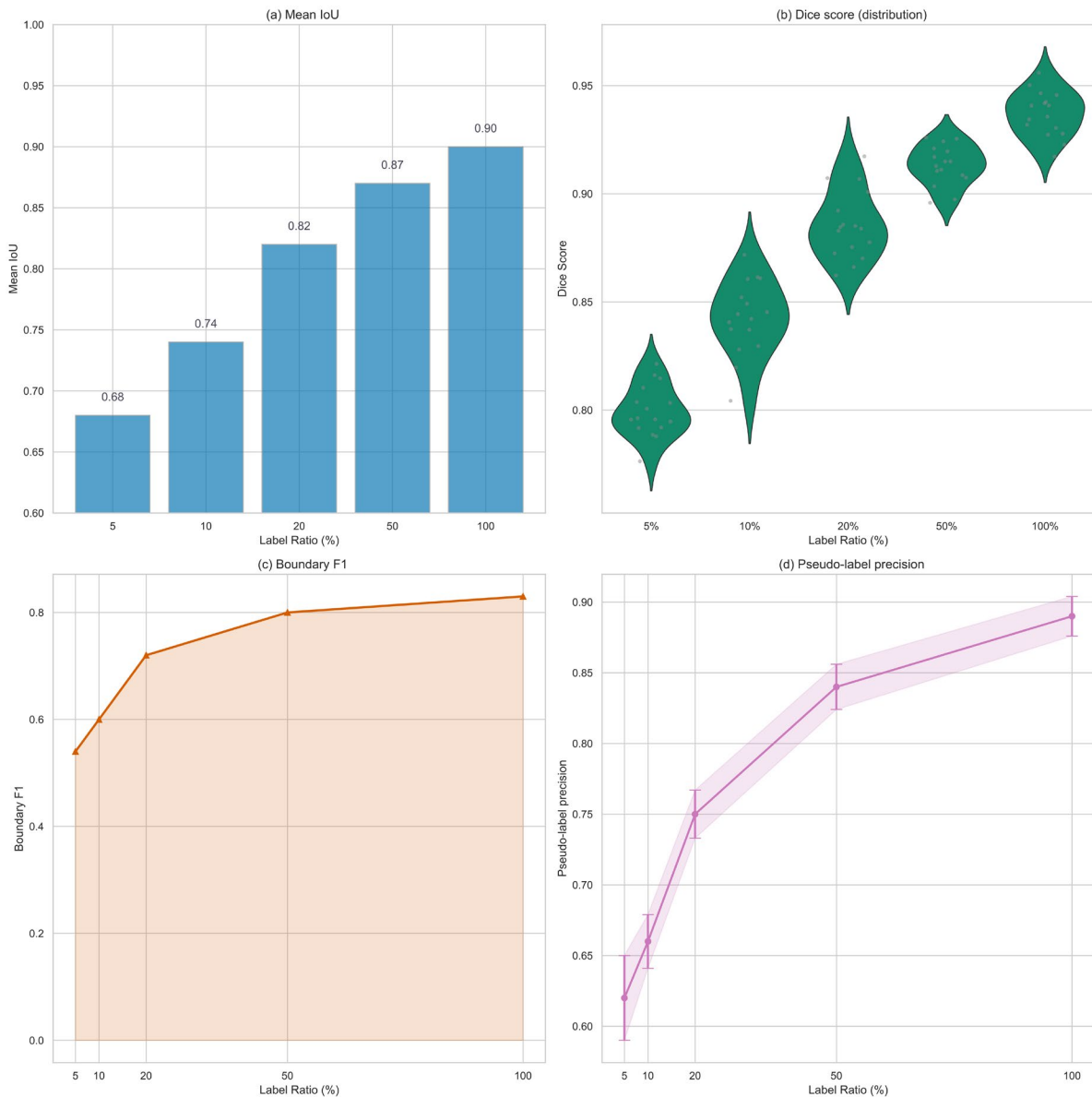


Figure 5. Segmentation performance under limited labels. (a) Mean IoU. (b) Dice score. (c) Boundary F1. (d) Pseudo-label precision.

Additionally, the four-performance metrics exhibit distinct model behaviors. At the lowest label setting, U-Net is more susceptible to sparse-label noise, but it improves quickly as the label ratio increases. Although Swin Transformer has a broad tissue context, fragmented marbling may occasionally be over-smoothed into a single spot. PatchMixer is a common patch mixer that improves local texture modeling; however, it is overly forgiving with ambiguous masks in the absence of pseudo-label filtering. Because the new model is more balanced, local texture swapping is added by the patch mixer, and the impact of unreliable unlabeled pixels is lessened by confidence filtering. Additionally, calibration research [20] demonstrates that not all high-confidence masks may be taken for granted and that probability dependability must be confirmed.

The border and texture-specific findings are displayed in Figure 6. The segmentation error for diffuse speckles, broad marbling bands, clustered fat islands, and narrow fat streaks is shown in Figure 6(a). Because multi-scale patch mixing connects local fragments without losing edge information, the model performs better with thin and clustered fat.

The marbling probability map's calibration inaccuracy is displayed in Figure 6(b). In reflecting and dark areas, the suggested confidence calibration lowers overconfident mistakes. The effectiveness of a segmentation probability is contingent upon its proximity to real accuracy, which is in line with the demands of contemporary

calibration analysis [21]. It will work especially well in the manufacturing setting when questionable masks are manually scrutinized. The operator should have some degree of confidence in the model's prediction that a pixel is likely to be marbling rather than treating it as a raw neural score.

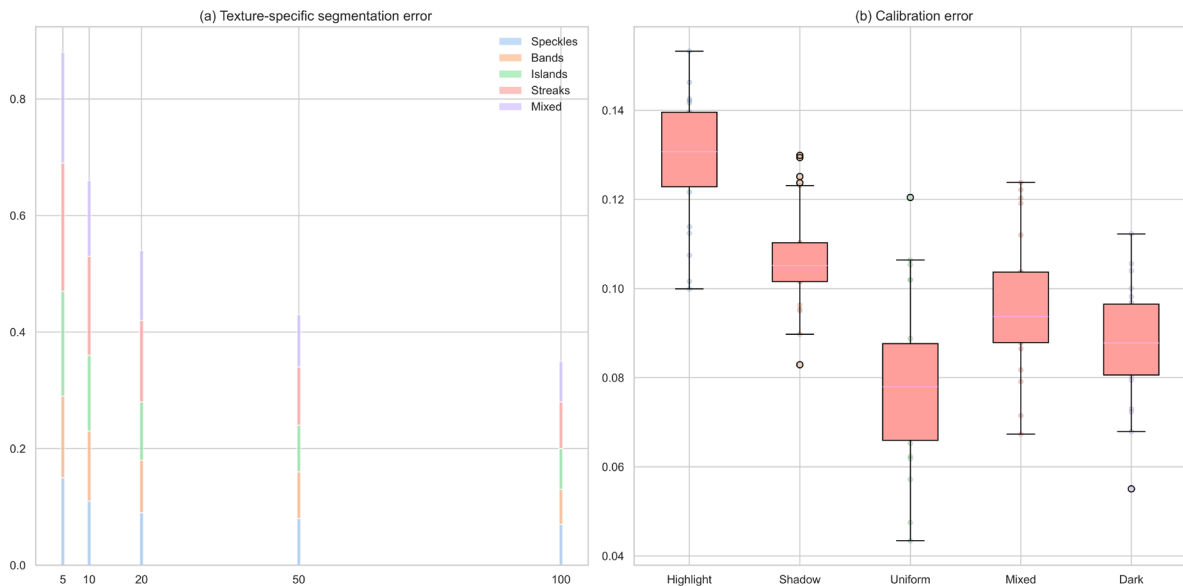


Figure 6. Boundary quality and probability reliability. (a) Texture-specific segmentation error. (b) Calibration error.

The study indicates that the suggested model is unlikely to result in specular reflections appearing as mottling. PatchMixer collects too many pixels that resemble fat and over-segments the bright knife marks. Thin streaks are not captured by U-Net, although having respectable coarse masks. Transformer-based models are capable of obtaining a rough contour of the tissue, but they might not be able to capture the fine features at the boundary between marbling and lean muscle. The suggested strategy retains the precise fat border and suppresses unreliable pseudo-labels by combining local patch information, multi-scale context, and confidence filtering. When forecasts are utilized for decision-making rather than just benchmark reporting, a kind of dependability layer is required, as explained by reviews on uncertainty quantification [22].

You might think of the segmentation curves as various meat grades. The Dice coefficient has a constant value across various marbling levels, while the Mean IoU indicates how effectively the algorithm recognizes the whole fat region. Boundary F1 is more susceptible to thin streaks, which can be observed as a slight shift in the marbling's fineness and consistency. The percentage of beneficial (or safe) samples among the unlabeled ones is known as pseudo-label precision. In the subsequent epochs, the pseudo-masks will have mistaken if the label budget is too low. Consequently, until the instructor branch has produced a stable probability map, the approach will not employ uncertain pixels.

Additionally, the border areas demonstrate that the training strategy and the model structure must be addressed together. Long-range dependencies in the tissue appearance can be learned by a strong encoder, but blurring at the fat-muscle interface will still happen in the absence of boundary-aware refinement. Simultaneously, a local model might maintain the sharp edge while failing to link the separate fat streaks that are part of the same marbling pattern. In order to learn both thin-streak continuity and large-scale fat distribution, PatchMixer facilitates both local interchange between adjacent patches and extended exchange at different scales. Thus, the operational danger of overconfident masks is decreased by the calibrating layer.

Additionally significant is the Pattern of Visual Error. Wet reflections, white connective tissue, and cut-surface scratches are the initial types of false alarms. Faint intramuscular fat that contrasts poorly with pale muscle is the primary cause of false negatives. The mistakes are recognized ambiguities in visual meat inspection; they are not random. The suggested model will display the ambiguity in a calibrated probability map, but it cannot resolve all ambiguous cases. As a result, a supervisor is not required to randomly wear a mask in an unclear region.

Model Ablation and Production-Line Applicability

The ablation and deployment analysis are displayed in Figure 7. The module-removal configuration without boundary loss, pseudo-label confidence filtering, multi-scale mixing, or calibration layer is shown in Figure 7(a). Low-quality pseudo-masks are harmful when labels are sparse because removing confidence filtering causes a comparatively greater decline in performance for the 5% label scenario.

The robustness under lighting variation, surface moisture, and cutting-texture disturbance is shown in Figure 7(b). Because it makes use of confidence-controlled pseudo-labels and texture-aware patch features, the suggested model is rather stable when compared to the baselines. Annotation saving and inference delay are shown in Figure 7(c). Although the model is heavier than a little U-Net, it is still practical for manufacturing lines.

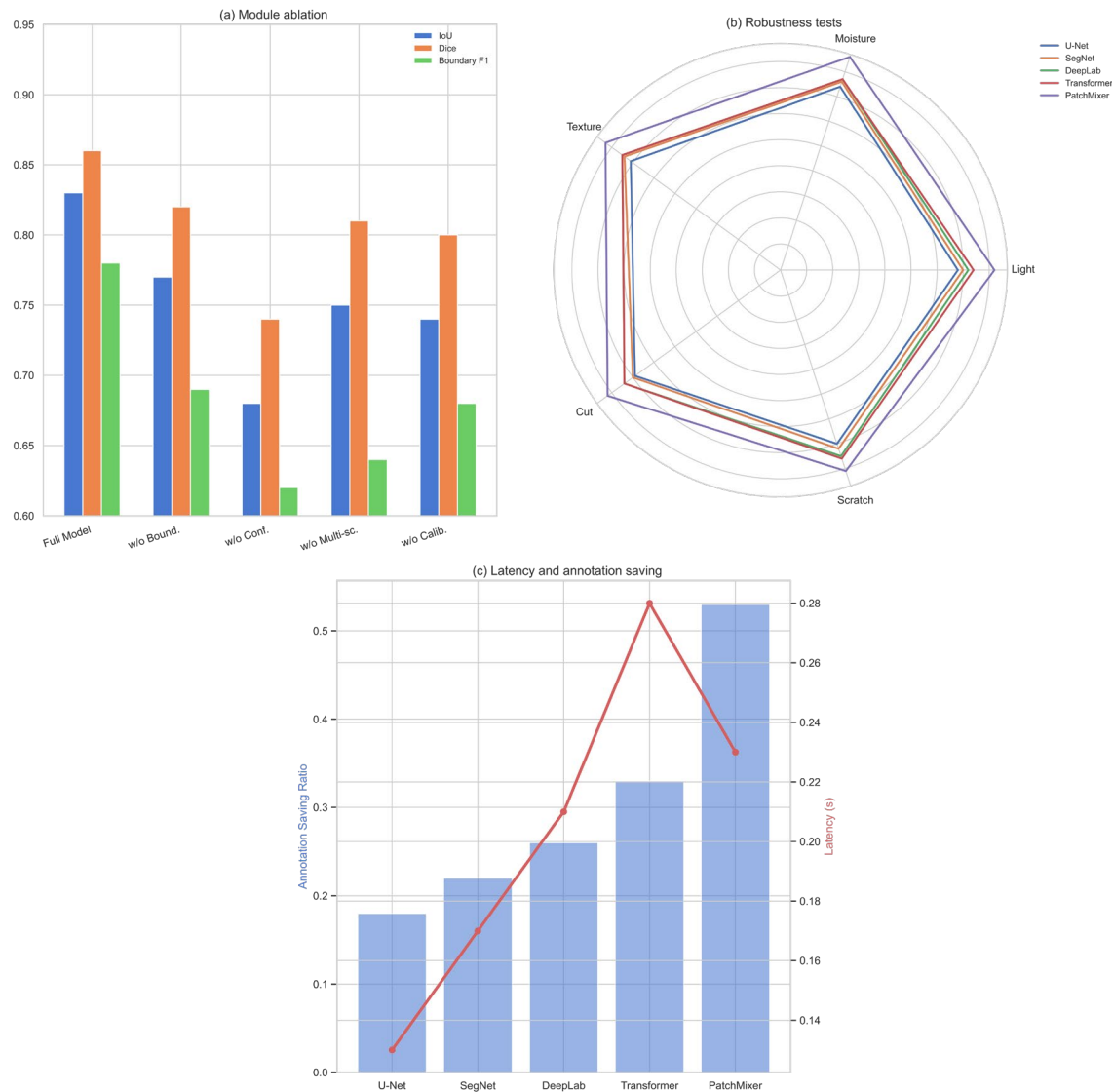


Figure 7. Ablation, robustness, and deployment value. (a) Module ablation. (b) Robustness tests. (c) Latency and annotation saving.

In reality, this means that not all expert ratings should be replaced by the model. A general-purpose partition may be made using it prior to grade estimate. The model may be used by the Production System to produce marbling masks, compute fat area ratios, pinpoint ambiguous areas, and send challenging photos for human review. Studies on visual quality inspection have shown a similar deployment strategy [23], and automation is only deemed practical when labor reduction does not jeopardize traceability. In this study, the operator may view the uncertain segmentation border using the confidence map instead of only receiving a binary response.

The model can lessen the need for sparse annotations. The teacher-student pseudo-label approach can increase the training data without thick masks for each batch if a plant can only identify a limited number of pictures. Manage the Pseudo-label Generation. Confidence filtering, boundary-aware loss, and calibration are all successful based on the ablation results. Working stability and division precision would not exist without any of the. A potential future expansion of the existing RGB-based procedure has been suggested since hyperspectral meat inspection research [24] has demonstrated that extra spectral evidence can enhance material classification. The concept of industrial inspection that combines imaging, prediction, and quality-control assessment is further supported by non-invasive meat authenticity work [25].

From an engineering perspective, the distinction between accuracy loss and workload reduction is the most useful ablation outcome. Eliminating pseudo-label filtering harms low-label training, while eliminating multi-scale mixing mostly harms fine-texture segmentation. Even though the region accuracy is quite high, boundary loss lowers edge quality and makes the mask less dependable without calibration. These failure modes correspond to various phases of an actual inspection procedure. For continuous operation, a manufacturing line must precisely split the region for marbling, have distinct limits in texture description, dependable uncertainty in review routing, and fast inference.

The model will be utilized as an intermediary perception module, according to deployment analysis. The downstream grading program computes quantitative descriptors including fat-area ratio, streak density, and texture continuity after the camera system first obtains a standardized meat-surface picture. The model then produces a marbling mask and an uncertainty map. If the uncertainty is concentrated in a limited area, the system will just flag that area for evaluation. The entire image will be marked for manual annotation or reacquisition if the uncertainty is too great. Because the model will not directly produce the finished commercial-grade product, a procedure is reasonably viable; otherwise, grading criteria differ by region and for different kinds of meat in accordance with the company's rules.

The remaining flaws will be dealt with sensibly. Since the study focuses on image-based marbling segmentation, chemical composition, tenderness, flavor, and eating quality cannot be directly determined. Although it is a broad indicator of meat quality, visual marbling does not account for every aspect. Segmentation instability can also result from extremely uneven light, extreme surface drying, ice, bloodstains, and packing film reflections. In this manner, it is also expected that there are sufficient right fat instances in the sparse labels for pseudo-label learning. The model won't be able to learn extremely fine marbling if the first labels only cover broad fat bands.

There are still certain shortcomings. True marbling may be concealed close to lean muscle fibers because low-contrast meat surfaces are hard to see. A pattern that resembles fat can also be produced by a strong specular reflection. Additionally, not all annotation behaviors in a real factory can be covered by simulated label budgets. Multispectral imaging, active learning, live annotation correction, and grade-level prediction following segmentation will all be included in future work. When appearance cues are not accessible, spectral and chemometric data can be added to visual segmentation, as demonstrated by fresh-beef hyperspectral prediction studies [26] and reviews on microbial-spoilage inspection [27]. To find out if the next iteration of the model can integrate marbling masks with various imaging techniques, additional studies have been conducted on lamb [28], pork composition [29], and red-meat hyperspectral systems [30]. The suggested model can advance from mask generation to complete meat quality decision support using a comparable integrated workflow.

Conclusion

This paper proposed a label-efficient PatchMixer model for segmenting meat marbling. A technique that reduces the need for dense pixel labeling by using confidence-guided learning, converting meat-surface pictures into overlapping patch tokens, and using multi-scale patch mixing to acquire fat streak and marbling cluster information. The results show that changing label ratios can improve the model's mean IoU, Dice score, boundary F1, and pseudo-label reliability over those attained using conventional segmentation baselines. Since the low-label situation shows the most gain, the approach could be more suited for industrial applications where expert mask annotation is expensive.

The first is that it may be employed in a label-efficient training mode and has comparatively excellent segmentation accuracy. Boundary-sensitive loss, probability calibration, consistency regularization, and pseudo-

label filtering are all integrated with sparse expert masks. For unlabeled meat photos, the modules assist the model in avoiding over-reliance on all pseudo-masks. The division findings may be used to calculate the fat area, examine the texture of the marbling, get ready for grading, and manually fix any unclear areas. The approach will maintain traceable visual evidence for quality control while reducing the amount of annotation effort in a production-line process.

Further data covering different types of meat, different cuts and grades, and multiple camera systems will be added in subsequent work. withstand extremely slight marbling, moisture fluctuations, and specular reflection more resiliently. By choosing just the most instructive photos for expert annotation, active learning can lower the annotation cost. It is easier to distinguish between surface artifacts and real intramuscular fat using multi-spectral or polarized pictures. To put it briefly, the label-efficient PatchMixer system that has been provided provides a workable foundation for high-scale, reliable, and financially viable beef marbling segmentation.

Author Contributions

Bogdan Radu contributes to conceptualization, methodology, software, validation, analysis, investigation, data collection, draft preparation, manuscript editing, visualization, supervision. Mihnea Lupu contributes to methodology, software, validation, analysis. All authors have read and agreed with the manuscript before its submission and publication.

Funding

This research received no specific financial support from any funding agency.

Institutional Review Board Statement

Not applicable.

References

- [1] Kakani, V. G., Nguyen, V. H., Kumar, B. P., Kim, H., & Pasupuleti, V. R. (2020). A critical review on computer vision and artificial intelligence in food industry. *Journal of Agriculture and Food Research*, 2, 100033. <https://doi.org/10.1016/j.jafr.2020.100033>
- [2] Antequera, T., Caballero, D., Grassi, S., Uttaro, B., & Perez-Palacios, T. (2021). Evaluation of fresh meat quality by hyperspectral imaging, nuclear magnetic resonance and magnetic resonance imaging: A review. *Meat Science*, 172, 108340. <https://doi.org/10.1016/j.meatsci.2020.108340>
- [3] Ouali, Y., Hudelot, C., & Tami, M. (2020). Semi-supervised semantic segmentation with cross-consistency training. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 12674-12684. <https://doi.org/10.1109/CVPR42600.2020.01269>
- [4] Chen, L. C., Zhu, Y., Papandreou, G., Schroff, F., & Adam, H. (2018). Encoder-decoder with atrous separable convolution for semantic image segmentation. *European Conference on Computer Vision*, 801-818. https://doi.org/10.1007/978-3-030-01234-2_49
- [5] Liu, Z., Lin, Y., Cao, Y., Hu, H., Wei, Y., Zhang, Z., et al. (2021). Swin Transformer: Hierarchical vision transformer using shifted windows. *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 10012-10022. <https://doi.org/10.1109/ICCV48922.2021.00986>
- [6] Zhu, M., Huang, D., Hu, X. J., Tong, W., Han, B., Tian, J., & Luo, H. (2020). Application of hyperspectral technology in detection of agricultural products and food: A review. *Food Science & Nutrition*, 8(10), 5206-5214. <https://doi.org/10.1002/fsn3.1852>
- [7] Adithya, S. C. (2022). A review of hyperspectral imaging in the quality evaluation of meat, fish, poultry and their products. *Meat Technology*, 63(2), 109-120. <https://doi.org/10.18485/meattech.2022.63.2.4>
- [8] Li, X., Yu, L., Chen, H., Fu, C. W., Xing, L., & Heng, P. A. (2021). Transformation-consistent self-ensembling model for semisupervised medical image segmentation. *IEEE Transactions on Neural Networks and Learning Systems*, 32(2), 523-534. <https://doi.org/10.1109/TNNLS.2020.2995319>
- [9] Chen, X., Yuan, Y., Zeng, G., & Wang, J. (2021). Semi-supervised semantic segmentation with cross pseudo supervision. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2613-2622. <https://doi.org/10.1109/CVPR46437.2021.00264>

- [10] Khan, S., Naseer, M., Hayat, M., Zamir, S. W., Khan, F. S., & Shah, M. (2022). Transformers in vision: A survey. *ACM Computing Surveys*, 54(10s), 1-41. <https://doi.org/10.1145/3505244>
- [11] Touvron, H., Cord, M., Jégou, H., & Bojanowski, P. (2022). ResMLP: Feedforward networks for image classification with data-efficient training. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 45(4), 5314-5321. <https://doi.org/10.1109/TPAMI.2022.3206148>
- [12] Tan, M., Pang, R., & Le, Q. V. (2020). EfficientDet: Scalable and efficient object detection. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 10781-10790. <https://doi.org/10.1109/CVPR42600.2020.01079>
- [13] Kervadec, H., Bouchtiba, J., Desrosiers, C., Granger, E., Dolz, J., & Ayed, I. B. (2021). Boundary loss for highly unbalanced segmentation. *Medical Image Analysis*, 67, 101851. <https://doi.org/10.1016/j.media.2020.101851>
- [14] Minaee, S., Boykov, Y., Porikli, F., Plaza, A., Kehtarnavaz, N., & Terzopoulos, D. (2022). Image segmentation using deep learning: A survey. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 44(7), 3523-3542. <https://doi.org/10.1109/TPAMI.2021.3059968>
- [15] Berman, M., Triki, A. R., & Blaschko, M. B. (2018). The Lovász-Softmax loss: A tractable surrogate for the optimization of the intersection-over-union measure in neural networks. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 4413-4421. <https://doi.org/10.1109/CVPR.2018.00464>
- [16] Sehar, U., & Naseem, M. L. (2022). How deep learning is empowering semantic segmentation: Traditional and deep learning techniques for semantic segmentation. *Multimedia Tools and Applications*, 81, 30519-30544. <https://doi.org/10.1007/s11042-022-12821-3>
- [17] Jadon, S. (2020). A survey of loss functions for semantic segmentation. *2020 IEEE Conference on Computational Intelligence in Bioinformatics and Computational Biology*, 1-7. <https://doi.org/10.1109/CIBCB48159.2020.9277638>
- [18] Ma, N., Zhang, X., Zheng, H. T., & Sun, J. (2018). ShuffleNet V2: Practical guidelines for efficient CNN architecture design. *European Conference on Computer Vision*, 122-138. https://doi.org/10.1007/978-3-030-01264-9_8
- [19] Zou, Y., Yu, Z., Liu, X., Kumar, B. V. K. V., & Wang, J. (2019). Confidence regularized self-training. *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 5982-5991. <https://doi.org/10.1109/ICCV.2019.00608>
- [20] Kull, M., Perello-Nieto, M., Kängsepp, M., Silva Filho, T., Song, H., & Flach, P. (2019). Beyond temperature scaling: Obtaining well-calibrated multiclass probabilities with Dirichlet calibration. *Advances in Neural Information Processing Systems*, 32, 12316-12326. <https://doi.org/10.5555/3454287.3455391>
- [21] Mukhoti, J., Kulharia, V., Sanyal, A., Golodetz, S., Torr, P. H. S., & Dokania, P. K. (2020). Calibrating deep neural networks using focal loss. *Advances in Neural Information Processing Systems*, 33, 15288-15299. <https://doi.org/10.5555/3495724.3497007>
- [22] Abdar, M., Pourpanah, F., Hussain, S., Rezazadegan, D., Liu, L., Ghavamzadeh, M., et al. (2021). A review of uncertainty quantification in deep learning: Techniques, applications and challenges. *Information Fusion*, 76, 243-297. <https://doi.org/10.1016/j.inffus.2021.05.008>
- [23] Ren, R., Hung, T., & Tan, K. C. (2018). A generic deep-learning-based approach for automated surface inspection. *IEEE Transactions on Cybernetics*, 48(3), 929-940. <https://doi.org/10.1109/TCYB.2017.2668395>
- [24] Shi, Y., Wang, X., Borhan, M. S., Young, J., Newman, D., Berg, E., & Sun, X. (2021). A review on meat quality evaluation methods based on non-destructive computer vision and artificial intelligence technologies. *Food Science of Animal Resources*, 41(4), 563-588. <https://doi.org/10.5851/kosfa.2021.e25>
- [25] Achata, E. M., Oliveira, M., Esquerre, C. A., Tiwari, B. K., & O'Donnell, C. P. (2020). Visible and NIR hyperspectral imaging and chemometrics for prediction of microbial quality of beef Longissimus dorsi muscle under simulated normal and abuse storage conditions. *LWT*, 128, 109463. <https://doi.org/10.1016/j.lwt.2020.109463>
- [26] Hitchman, S. C., Loeffen, M. P. F., Reis, M. M., & Craigie, C. R. (2021). Robustness of hyperspectral imaging and PLSR model predictions of intramuscular fat in lamb M. longissimus lumborum across several flocks and years. *Meat Science*, 179, 108492. <https://doi.org/10.1016/j.meatsci.2021.108492>
- [27] Odeyemi, O. A., Alegbeleye, O. O., Strateva, M., & Stratev, D. (2020). Understanding spoilage microbial community and spoilage mechanisms in foods of animal origin. *Comprehensive Reviews in Food Science and Food Safety*, 19(2), 311-331. <https://doi.org/10.1111/1541-4337.12526>

- [28] Wang, C., Wang, S., He, X., Wu, L., Li, Y., & Guo, J. (2020). Combination of spectra and texture data of hyperspectral imaging for prediction and visualization of palmitic acid and oleic acid contents in lamb meat. *Meat Science*, 169, 108194. <https://doi.org/10.1016/j.meatsci.2020.108194>
- [29] Rady, A., & Adedeji, A. A. (2020). Application of hyperspectral imaging and machine learning methods to detect and quantify adulterants in minced meats. *Food Analytical Methods*, 13, 970-981. <https://doi.org/10.1007/s12161-020-01719-1>
- [30] Uddin, M. N. (2021). Impact of near infrared spectroscopy and hyperspectral imaging system to predict physicochemical composition and quality attributes of meat: A review. *Journal of Agriculture, Food and Environment*, 2(3), 66-73. <https://doi.org/10.47440/JAFE.2021.2310>