

Article

Automatic Assessment of Fabric Soil Release Appearance Using a Lightweight Tri-Semantic Injection Network with Ordinal Learning

Wen-Yang Chang ¹, Cheng-Hsun Huang ^{2,3,*}  and Li-Wei Chen ¹ 

¹ Department of Mechanical and Computer-Aided Engineering, National Formosa University, Yunlin 632301, Taiwan; wenyang@nfu.edu.tw (W.-Y.C.); liwei@nfu.edu.tw (L.-W.C.)

² Doctoral Degree Program in Smart Industry Technology Research and Development, College of Engineering, National Formosa University, Yunlin 632301, Taiwan

³ Taiwan Textile Research Institute, New Taipei City 23674, Taiwan

* Correspondence: d1381101@nfu.edu.tw; Tel.: +886-939-202-536

Abstract

Soil release appearance grading evaluates residual stains after standardized laundering, but visual assessment is subjective and adjacent half grades are difficult to distinguish. This study proposes a lightweight Tri-Semantic Injection Network (TSI-Net) for nine-grade assessment. From one red–green–blue (RGB) image, a fixed CIE L*a*b* (CIELAB) branch constructs a mean-background image B, a pixel-wise color-difference image D, and a stain-appearance image S. The trainable backbone progressively injects S, D, and B and predicts grades from 1.0 to 5.0 at 0.5-grade intervals. Gaussian soft targets represent the ordering of these grades. The dataset contains 325 images, and TSI-Net has 1,403,336 trainable parameters. In a 33-image evaluation, Gaussian-trained TSI-Net achieved exact-grade and within-half-grade accuracies of 87.88% and 96.97%, compared with 84.85% and 93.94% for one-hot training. Both training methods achieved 100.00% accuracy within one grade. Gaussian training therefore showed a 3.03-percentage-point advantage for each of the two stricter metrics in this comparison. This method requires no manual stain segmentation and provides a compact framework for standardized fabric appearance assessment under controlled acquisition conditions.

Keywords: fabric soil release; industrial machine vision; textile quality inspection; lightweight neural network; ordinal learning



Academic Editor: Dimitris Mourtzis

Received: 24 August 2026

Revised: 7 September 2026

Accepted: 9 September 2026

Published: 11 September 2026

Copyright: © 2026 by the authors.

Licensee MDPI, Basel, Switzerland.

This article is an open access article distributed under the terms and conditions of the [Creative Commons Attribution \(CC BY\)](https://creativecommons.org/licenses/by/4.0/) license.

1. Introduction

Soil release performance is an important functional property of textile materials used in apparel, household products, medical textiles, and industrial applications. After a fabric is exposed to household soils and subsequently laundered, the visibility of the residual stain directly affects its perceived cleanliness, serviceability, and product quality. In textile laboratories and production quality control workflows, reliable assessment of soil release appearance is therefore essential for material development, performance verification, product release, and process traceability.

AATCC (American Association of Textile Chemists and Colorists) TM215-2023 [1] provides a standardized method for evaluating the ability of textile materials to release household soils during laundering. After the prescribed staining and laundering procedures, the residual stain appearance is visually compared with the AATCC Stain Release

Replica under controlled viewing conditions. The appearance is graded from Grade 1 to Grade 5, where Grade 1 represents the most visible residual staining and the poorest soil release performance, whereas Grade 5 represents the least visible residual staining and the best performance. In the present dataset, intermediate assessments were recorded at 0.5-grade intervals when the residual stain appearance fell between two adjacent grades. Consequently, nine ordered labels, ranging from Grade 1 to Grade 5, were included in the image-based assessment task.

For readers outside textile testing, the task is not to locate a discrete defect or identify the soil type. It is to reproduce the single overall appearance grade assigned by a trained assessor after laundering. The intended use is decision support: the model should help laboratories and production personnel apply the ordered grading scale more consistently, especially when two neighboring half grades appear similar.

Although visual grading is widely adopted in textile testing, its consistency can be influenced by assessor experience, illumination conditions, observation angle, and subjective interpretation. These effects are particularly evident when adjacent grades exhibit only subtle differences in stain intensity, boundary definition, or affected area. Manual assessment also becomes time-consuming when large numbers of specimens must be evaluated. An automated machine-vision system could improve inspection efficiency, repeatability, and traceability while providing quantitative decision support for trained inspectors.

Machine vision has been increasingly applied to textile appearance evaluation and industrial surface quality inspection. Digital imaging methods have been used to evaluate textile color fastness and staining by extracting colorimetric, histogram-based, and texture-related features [2–5]. For color-aware grading, Liang et al. [2] used spectral reconstruction and a backpropagation neural network to predict color-fastness-to-rubbing grades from the CIEDE2000 color difference and differences in CIELAB attributes (ΔL^* , Δa^* , and Δb^*). Shao et al. demonstrated that residual stain appearance could be quantified using image histogram analysis [6]. Peng et al. developed an improved YOLOv8 model for leather defect detection [7]. Defect localization differs from soil release assessment, which requires an overall ordered grade for residual household stains after laundering.

Image-based methods developed for other textile appearance properties provide an important foundation for automatic grading. Wang et al. formulated fabric smoothness evaluation as an ordinal classification task and developed a compact convolutional neural network with adjacent-label smoothing [8]. Their study incorporated grade order into fabric smoothness assessment. More recently, Shi et al. introduced conditional ordinal regression for neural networks (CORN), which uses conditional probabilities to obtain rank-consistent cumulative probabilities [9]. These studies illustrate different ways to incorporate grade order into learning, while differing in their target formulations and application domains. Fabric smoothness primarily concerns wrinkles and surface deformation, whereas soil release grading depends on residual stain color, intensity, area, distribution, and contrast with the surrounding clean fabric.

Recent lightweight vision architectures offer strategies for computationally efficient visual recognition. MobileNetV4 combines efficient building blocks with neural architecture search to support recognition across mobile hardware [10]. These advances address computational efficiency and ordered prediction, while the soil release task additionally requires residual stain appearance to be interpreted relative to the surrounding fabric. The present study addresses this task-specific requirement through a compact RGB network with Gaussian ordinal supervision and progressive injection of CIELAB-derived stain appearance, difference, and background representations.

Automatic soil release appearance grading remains challenging for several reasons. First, different household soils may produce substantial variations in color, shape, size,

and spatial distribution. Second, the perceived severity of a residual stain depends not only on the stained region but also on its chromatic difference from the surrounding fabric. Third, fabric composition, inherent color, texture, and illumination variation may interfere with stain recognition. Fourth, the visual differences between adjacent grades, particularly half-grade intervals, can be subtle. Finally, conventional one-hot classification assigns zero target weight to every non-reference grade, even though an adjacent half grade is more similar to the reference than a distant grade. A practical industrial solution must address these issues while maintaining a compact model suitable for laboratory and production use.

To address these challenges, this study proposes a lightweight Tri-Semantic Injection Network, referred to as TSI-Net, for machine-vision-based soil release appearance assessment. From a single RGB image, a fixed procedure in the CIE L*a*b* (CIELAB) color space calculates a per-image mean background color from the outer border band and produces three complementary semantic images: stain appearance *S*, stain-to-background color difference *D*, and mean background *B*. In plain terms, *B* represents what the surrounding clean fabric looks like, *D* measures how different each pixel is from that reference, and *S* retains the weighted color and spatial pattern of the residual stain. These images are injected progressively into the trainable RGB pathway in the order $S \rightarrow D \rightarrow B$. The classifier directly uses the pooled 256-dimensional visual representation and predicts nine ordered grades from 1.0 to 5.0 at intervals of 0.5.

Because soil release appearance labels possess an inherent order, TSI-Net uses Gaussian ordinal soft-label learning to assign greater target weight to grades nearer the reference grade. The evaluation metrics are exact-grade accuracy (Acc0), within-half-grade accuracy (Acc0.5), within-one-grade accuracy (Acc1), and trainable parameter count.

The principal contributions of this study are summarized as follows:

1. A machine-vision dataset comprising 325 reference-graded fabric images was established. It includes 100% cotton, 100% polyester, and 65% polyester/35% cotton fabrics and covers nine ordered grades from Grade 1 to Grade 5 at 0.5-grade intervals.
2. A lightweight TSI-Net architecture was developed to inject a three-channel stain appearance image *S*, a one-channel CIELAB difference image *D*, and a three-channel mean-background image *B* progressively into three convolutional stages.
3. Gaussian ordinal soft-label learning was incorporated into the proposed tri-semantic framework to represent the ordered relationship among the nine soil release appearance grades by assigning larger target probabilities to grades closer to the reference grade.
4. The comparison included AlexNet, ResNet18, ResNet34, InceptionV3, and VGG16, together with five semantic-input configurations. Sections 5.1 and 5.2 report the results obtained on the common 33-image held-out test set. TSI-Net achieved 87.88% Acc0, 96.97% Acc0.5, and 100.00% Acc1, with 1,403,336 trainable parameters. A separate comparison of TSI-Net trained with one-hot and Gaussian soft targets is reported in Section 5.3.

2. Dataset Preparation

2.1. Image Acquisition and Reference Grading

The fabric images used in this study were obtained from a professional textile testing unit at the Taiwan Textile Research Institute. The specimens covered three common fiber compositions: 100% cotton, 100% polyester, and 65% polyester/35% cotton. The dataset contained 105 images of 100% cotton, 110 images of 100% polyester, and 110 images of 65% polyester/35% cotton, for a total of 325 fabric images with individual reference grades.

One photograph was acquired for each physical specimen; multiple photographs of the same specimen were not included.

Specimen preparation and soil release assessment followed AATCC TM215-2023 [1]. The soil categories recorded in available source-image file names included chili sauce, mustard, tomato sauce, mayonnaise, coffee, soy sauce, and blood. Laundering used the standard normal cycle specified in AATCC LP1-2021, with a wash-water temperature of 41 °C. Specimens and laundering ballast formed a total load of 1.8 ± 0.1 kg, and 66 ± 1 g of AATCC 1993 Standard Reference Detergent [11] was added.

After washing, tangled specimens and ballast were separated with minimal distortion. The load was tumble-dried until fully dry, using a maximum exhaust temperature of 68 ± 6 °C. Specimens were removed immediately after drying and laid flat to prevent wrinkles and creases that could affect grading. Soil release appearance was evaluated within 4 h after drying; reconditioning was not required.

Image acquisition was conducted in the textile testing laboratory at the Taiwan Textile Research Institute under controlled laboratory conditions. A CCD camera fitted with a camera lens was mounted on a copy stand. Each specimen was laid flat on the baseboard, with the camera optical axis perpendicular to the specimen plane (90° to the plane). The field of view included the residual stain and surrounding unstained fabric, preserving both stain appearance and contrast with the background.

The exposure time was set to 10 ms, the white-balance color temperature setting to 5500 K, the gain to 0 dB, and gamma to 1.0. The lens aperture was fixed at f/5.6. Illumination was provided by the standard light source prescribed by the testing laboratory for fabric appearance inspection. The lighting arrangement and ambient conditions were kept unchanged for all specimens, and the setup was positioned so that the camera, lens, and copy stand did not cast visible shadows within the ROI. These fixed imaging conditions were used throughout data acquisition to maintain consistent image brightness and color. Figure 1 illustrates the normal-view acquisition geometry.

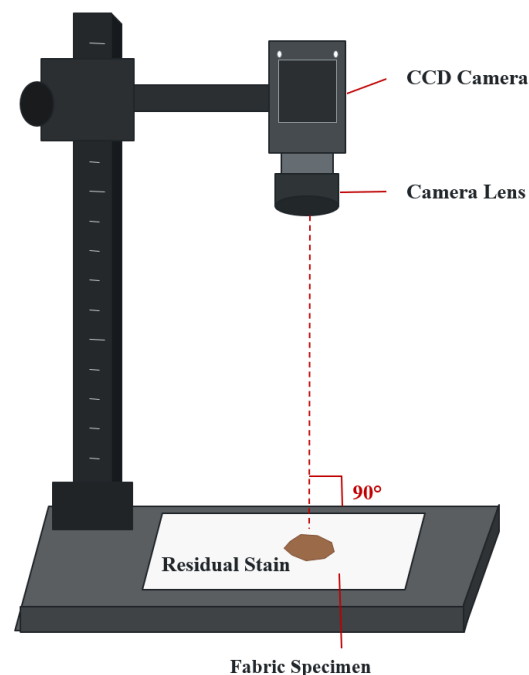


Figure 1. Schematic of the image acquisition setup under the standard laboratory illumination prescribed for fabric appearance inspection. A CCD camera with a camera lens was mounted on a copy stand above a flat fabric specimen. The optical axis was perpendicular to the specimen plane (90°). The schematic is not to scale.

Figure 2 provides representative examples of the visible grading task at the five integer anchor grades. The examples illustrate the type of residual stain appearance evaluated by trained assessors; the intermediate half grades used for model development and evaluation are not shown.

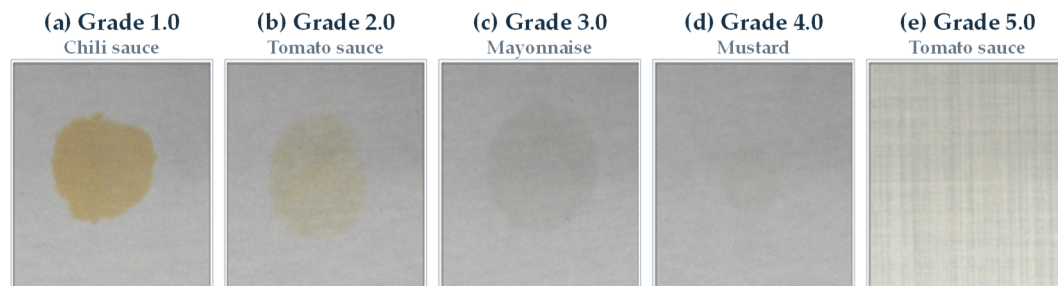


Figure 2. Representative experimental images at the five integer anchor grades.

Across these examples, the residual stain generally becomes less visually distinct from Grade 1.0 to Grade 5.0, which is consistent with the grading definition. However, the panels show different specimens and soil types rather than a controlled progression of one specimen. Figure 2 is therefore illustrative and should not be used to infer a monotonic relationship between grade and any single attribute such as stain color or area.

2.2. Data Grading and Dataset Composition

Reference grading was performed by two professionally trained assessors at the Taiwan Textile Research Institute. The specimens were divided between the assessors, and each specimen received one grade from its assigned assessor. No specimen was independently rated by both assessors; therefore, no averaging or disagreement resolution procedure was applied. Each assessor compared the residual stain appearance with the AATCC Stain Release Replica in accordance with AATCC TM215-2023 under controlled laboratory viewing conditions. Grade 1 denotes the most visible residual staining and the poorest soil release appearance, whereas Grade 5 denotes the least visible residual staining and the best appearance. A half-grade label was recorded when the appearance fell between adjacent integer grades, giving nine ordered labels: 1, 1.5, 2, 2.5, 3, 3.5, 4, 4.5, and 5.

Each image received one global appearance label; no bounding box annotation was required because the objective was specimen-level grading rather than defect localization. The 325 images were distributed across three fiber compositions and nine grades, as shown in Table 1. Grade counts ranged from 32 to 39, and the three fiber groups contained 105, 110, and 110 images, respectively. Thus, no grade or fiber group dominated the dataset. Figure 3 shows how the 325 images are distributed by grade and fiber composition. Knitted and woven fabrics were represented in an approximately 3:1 ratio (75% and 25%, respectively). No finishing treatment was applied to the test fabrics. All 325 images were acquired in a single imaging batch under the conditions described in Section 2.1.

Table 1. Distribution of fabric images according to fiber composition and soil release appearance grade.

Fiber Composition	1	1.5	2	2.5	3	3.5	4	4.5	5	Total
100% Cotton (C)	17	8	14	9	11	16	10	12	8	105
100% Polyester (T)	9	16	7	18	12	8	15	11	14	110
65% Polyester/35% Cotton (65% T/35% C)	13	10	17	8	15	12	9	16	10	110
Total	39	34	38	35	38	36	34	39	32	325

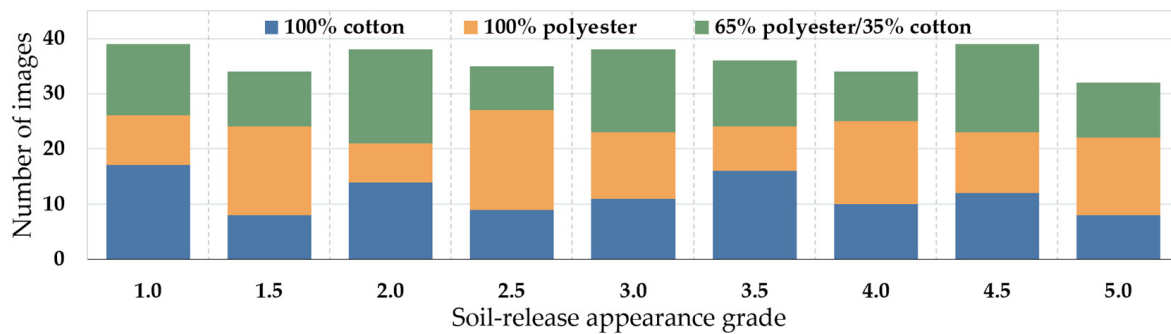


Figure 3. Grade-wise distribution of the 325 images. Each stacked bar gives the total number of images at one grade and partitions that total by fiber composition.

The dataset consisted of 325 images across 7 soil types. Mustard, tomato sauce, and mayonnaise each accounted for 47 images (14.46%), while chili sauce, soy sauce, blood, and coffee each accounted for 46 images (14.15%). The seven soil types were therefore nearly equally represented.

Visual categorization of the unstained fabric within the ROIs identified 124 images with white or off-white backgrounds (38.15%), 63 images with gray or light gray backgrounds (19.38%), and 46 images each with beige, light blue, and blue backgrounds (14.15% per category). These categories describe the fabric background appearance in the acquired images.

2.3. Image Preprocessing

Images were decoded as eight-bit RGB and interpreted as sRGB by the preprocessing implementation. The trainable pathway used 224×224 inputs with pixel intensities scaled to $[0, 1]$, whereas the fixed semantic branch used a 112×112 copy to generate B ($112 \times 112 \times 3$), D ($112 \times 112 \times 1$), and S ($112 \times 112 \times 3$) in the fixed order $B \rightarrow D \rightarrow S$. The RGB conversion and CIELAB channel scaling are defined in Section 3.2. No manual stain segmentation or additional input image was required.

3. Proposed TSI-Net

The purpose of TSI-Net is to estimate the specimen-level ordered grade used in standardized soil release assessment from a single RGB image while keeping the visual evidence understandable. Unlike a defect detector, the model does not output a bounding box or binary stain mask; it predicts one overall appearance grade. TSI-Net consists of four coordinated components: (1) fixed CIELAB semantic image extraction, (2) a lightweight RGB feature extraction backbone, (3) three-stage progressive semantic injection, and (4) nine-grade prediction using Gaussian soft-target cross-entropy. In plain terms, B represents the surrounding fabric background, D quantifies color separation from that background, and S represents the weighted stain appearance. The fixed branch provides these explicit cues, whereas the trainable branch learns the features needed to distinguish neighboring grades. Figure 4 qualitatively illustrates representative RGB and semantic images; it is not a measurement of classification accuracy.

In the selected examples, S retains the spatial distribution and color characteristics of the visible residual stain, D highlights pixels that differ from the specimen-specific mean background, and B represents the surrounding fabric color. Figure 4 presents representative semantic images, Figure 5 identifies the trainable injection points through which they enter the RGB pathway before classification, and Figure 6 illustrates how the stain- and background-related representations are constructed. Section 5.2 reports the performance of the evaluated semantic-input configurations. Because the configurations differ in trainable

parameter count, the comparison provides configuration-level evidence rather than a parameter-matched estimate of the isolated effect of any single semantic input.

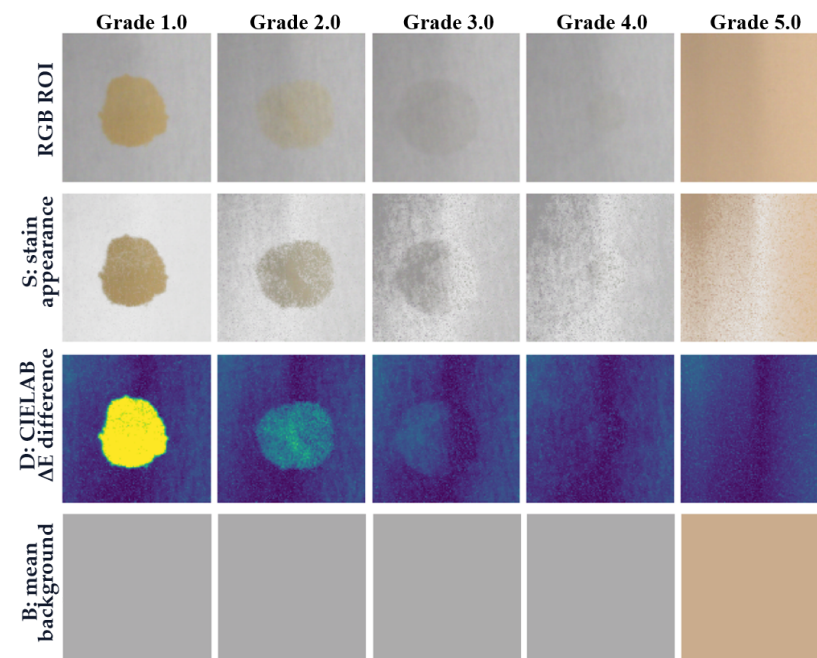


Figure 4. Qualitative examples of the simplified semantic images at the five integer anchor grades. Rows show the RGB ROI, stain appearance image S , CIELAB difference image D , and mean background image B ; columns correspond to different specimens labeled Grades 1.0, 2.0, 3.0, 4.0, and 5.0. For visualization only, D is displayed in false color, and RGB renderings are used for S and B ; the network receives the corresponding numerical tensors.

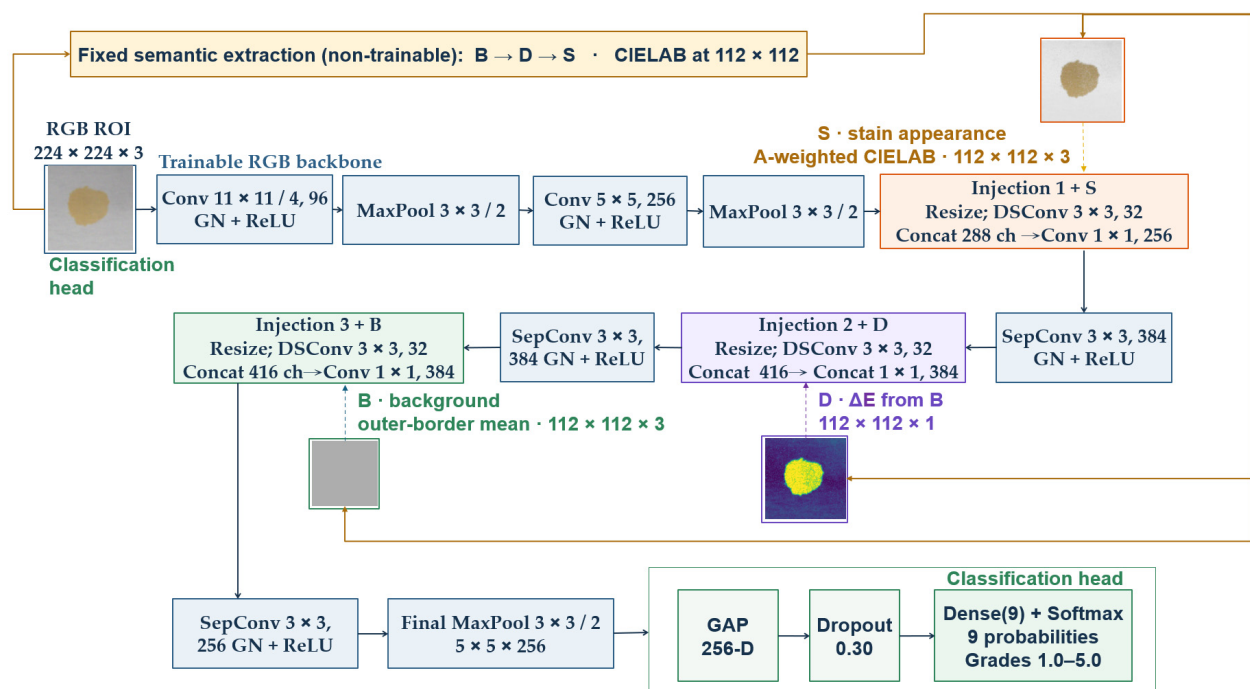


Figure 5. Detailed architecture of the nine-grade TSI-Net. The fixed branch extracts B , D , and S from the same RGB ROI in the order $B \rightarrow D \rightarrow S$. The trainable backbone progressively injects S , D , and B through resizing, depthwise-separable encoding, channel concatenation, and 1×1 fusion. The final $5 \times 5 \times 256$ backbone output is globally averaged to 256 dimensions, followed by Dropout(0.30) and Dense(9) with Softmax.

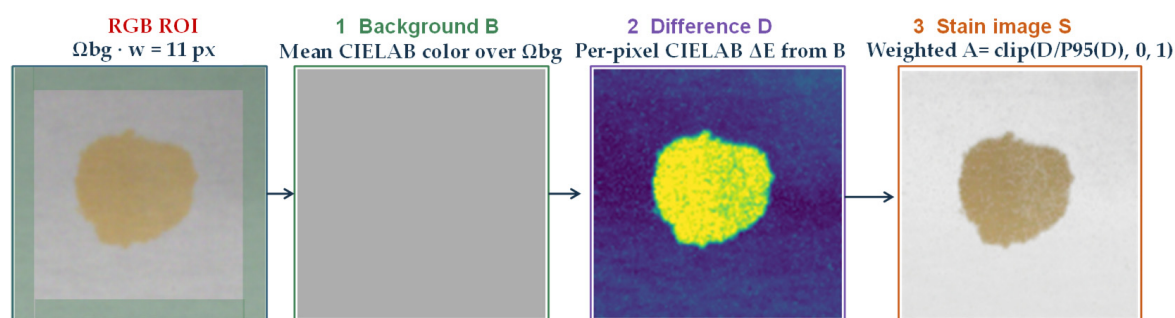


Figure 6. Extraction from one image. The outer-band mean forms B; its CIELAB distance forms D; D is normalized to form A; and A weights the normalized CIELAB image to form S. Extraction follows $B \rightarrow D \rightarrow S$, whereas injection follows $S \rightarrow D \rightarrow B$.

3.1. Overall Architecture of TSI-Net

TSI-Net receives a single RGB fabric image with dimensions $224 \times 224 \times 3$. The fixed semantic branch resizes the same image to 112×112 pixels and extracts B, D, and S in the fixed order $B \rightarrow D \rightarrow S$. Their tensor dimensions are $B, S \in \mathbb{R}^{112 \times 112 \times 3}$; $D \in \mathbb{R}^{112 \times 112 \times 1}$. In each tensor size, the three numbers denote height, width, and channel count, respectively. In parallel, the trainable RGB pathway reduces the input to a $12 \times 12 \times 256$ feature tensor and progressively injects S, D, and B in the order $S \rightarrow D \rightarrow B$. The final $5 \times 5 \times 256$ tensor is transformed by global average pooling into a 256-dimensional vector, followed by Dropout (0.30) and Dense(9) with Softmax. During training, Dropout(0.30) randomly sets 30% of the elements in the pooled feature vector to zero. Dense(9) is a fully connected output layer with nine neurons, one for each grade. Table 2 lists the principal operations and output sizes. The detailed architecture and the progressive $S \rightarrow D \rightarrow B$ injection pathway are illustrated in Figure 5.

Table 2. Principal layers, operations, and output sizes of the nine-grade TSI-Net.

Stage	Operation	Output Size
Input	RGB ROI	$224 \times 224 \times 3$
Fixed semantic branch	Resize to 112×112 ; convert to CIELAB; form B from the outer-band mean; calculate unscaled D; form A by percentile scaling and clipping; scale the CIELAB channels to construct S and the background input (Equation (3))	S: $112 \times 112 \times 3$;
		D: $112 \times 112 \times 1$;
		B: $112 \times 112 \times 3$
Stem 1	Conv $11 \times 11 / 4, 96$; GN; ReLU	$54 \times 54 \times 96$
Pool 1	MaxPool $3 \times 3 / 2$	$26 \times 26 \times 96$
Stem 2	Conv $5 \times 5, 256$; GN; ReLU	$26 \times 26 \times 256$
Pool 2	MaxPool $3 \times 3 / 2$	$12 \times 12 \times 256$
Injection 1	Resize S; DSConv $3 \times 3, 32$; GN+ReLU; concatenate; Conv $1 \times 1, 256$; GN+ReLU	$12 \times 12 \times 256$
Refinement 1	SepConv $3 \times 3, 384$; GN; ReLU	$12 \times 12 \times 384$
Injection 2	Resize D; DSConv $3 \times 3, 32$; GN+ReLU; concatenate; Conv $1 \times 1, 384$; GN+ReLU	$12 \times 12 \times 384$
Refinement 2	SepConv $3 \times 3, 384$; GN; ReLU	$12 \times 12 \times 384$
Injection 3	Resize B; DSConv $3 \times 3, 32$; GN+ReLU; concatenate; Conv $1 \times 1, 384$; GN+ReLU	$12 \times 12 \times 384$
Refinement 3	SepConv $3 \times 3, 256$; GN; ReLU	$12 \times 12 \times 256$
Final pool	MaxPool $3 \times 3 / 2$	$5 \times 5 \times 256$

Table 2. *Cont.*

Stage	Operation	Output Size
GAP	Global average pooling	256
Dropout	Dropout, rate = 0.30	256
Classifier	Dense(9) + Softmax; grades 1.0–5.0 at 0.5-grade intervals	9 probabilities

Figure 5 is read from the RGB input along the upper row from left to right, then along the middle row from right to left, and finally along the bottom row from left to right toward the classification head. Blue blocks represent the trainable RGB feature pathway, whereas orange, purple, and green blocks denote the S, D, and B injection stages, respectively. Tensor dimensions are reported as height $\times$ width $\times$ channels, and the upper CIELAB branch is fixed and contains no trainable parameters. For legibility, the injection block labels omit GN and ReLU; the complete operations are specified in Table 2 and Equation (4).

The RGB stem reduces the input to $12 \times 12 \times 256$ before semantic injection. At each stage, the selected semantic image is resized, encoded into 32 channels, joined with the current RGB feature tensor, and fused by a 1×1 convolution. Separate SepConv layers then refine the representation to 384, 384, and 256 channels. Independent TensorFlow/Keras reconstruction confirmed 1,403,336 trainable parameters.

The diagram makes two design choices explicit: the same RGB ROI supplies both branches, and the semantic tensors are resized and fused at three intermediate stages rather than appended only at the classifier. This keeps the final pooled representation compact and associates each cue with a defined injection point.

3.2. Simplified Mean-Referenced Semantic-Image Extraction

The fixed branch describes a residual stain relative to the clean fabric surrounding it, because the perceived soil release grade depends on both the stain and its contrast with the fabric background. Starting from the same RGB region of interest (ROI), the branch performs four connected tasks: (a) it estimates a specimen-specific mean background image B from the outer border; (b) it calculates a one-channel CIELAB difference image D and uses it to construct a three-channel stain appearance image S; (c) it defines the semantic role and tensor dimension of each image; and (d) it distinguishes the dependency-driven extraction order $B \rightarrow D \rightarrow S$ from the network injection order $S \rightarrow D \rightarrow B$. All operations are deterministic and contain no trainable parameters. B, D, and S are generated at 112×112 pixels, with three, one, and three channels, respectively. Figure 6 first presents this extraction logic visually; the equations then formalize the same steps.

In this example, the clean border yields a uniform reference B, while D and S emphasize the visible central stain without a manually drawn contour. Together with Figure 5, which traces S, D, and B through the three trainable injection stages, the visualization explains how the extracted semantic representations reach the classification pathway. The corresponding configuration-level performance is reported in Section 5.2. The diagram uses a compact weight expression; Equation (3) specifies the positive denominator floor and the CIELAB channel scaling explicitly. The reliability of the extractor depends on the outer band being representative of clean fabric. Stain contact with the border, strong illumination gradients, or patterned backgrounds can bias B and produce non-stain responses in D and S.

The cropped ROI was decoded using the Pillow library in RGB channel order. A 224×224 image was generated by bilinear interpolation for the RGB input, and a 112×112 copy was generated for constructing the semantic representations. The 8-bit RGB values were converted to float32 and divided by 255 to obtain encoded sRGB values in the range [0, 1]. The encoded sRGB image was converted to CIELAB using the standard

implementation provided by the scikit-image library, with the D65 illuminant and the 2° standard observer. The unscaled CIELAB values resulting from this conversion were used in Equations (1) and (2). Before being introduced into the network, the semantic channels were scaled according to Equation (3).

(a) Mean Background Image B. Residual stain visibility must be judged relative to the surrounding clean fabric, and fabric color or illumination may vary among specimens. The outer border is therefore used to estimate a specimen-specific reference without manually selecting a clean region. Equation (1) presents the core mean-and-repeat operation first; the following lines define the border width and border pixel set used to calculate that mean.

$$\begin{aligned}
 B_c(x, y) &= \bar{b}_c, \quad c \in \{L^*, a^*, b^*\} \\
 \bar{b}_c &= \frac{1}{|\Omega_{bg}|} \sum_{(u,v) \in \Omega_{bg}} I_c(u, v) \\
 w &= \lfloor 0.10 \min(H, W) \rfloor \\
 \Omega_{bg} &= \{(u, v) \mid 0 \leq u < W, 0 \leq v < H, \\
 &\quad u < w \vee u \geq W - w \vee v < w \vee v \geq H - w\}
 \end{aligned} \tag{1}$$

In Equation (1), the RGB ROI is resized to 112×112 pixels and converted to CIELAB [12], where L^* represents lightness, a^* represents the green–red axis, and b^* represents the blue–yellow axis. $B_c(x, y)$ is the value of B at (x, y) ; $\bar{b}_c$ is the mean of channel $c \in \{L^*, a^*, b^*\}$ over the border set Ω_{bg} ; (u, v) indexes a pixel in Ω_{bg} ; and $I_c(u, v)$ is its channel value. H and W are image height and width, $w = \lfloor 0.10 \min(H, W) \rfloor$ is the integer border width, and $\lfloor \cdot \rfloor$ denotes rounding downward. In the set definition, the vertical bar means “such that”, $\vee$ denotes logical OR, and $|\Omega_{bg}|$ is the number of border pixels. For $H = W = 112$, $w = 11$ pixels. Thus, Ω_{bg} is the complete outer band rather than only the four corners, and B is a uniform $112 \times 112 \times 3$ reference image. This simple reference depends on the border being sufficiently clean and representative.

(b) Difference Image D and Stain-Appearance Image S. Absolute pixel color alone is not a reliable measure of stain visibility across fabrics with different base colors. The image D therefore measures each pixel relative to the specimen-specific background, whereas S preserves weighted color and spatial information for the residual stain.

After the background reference has been obtained, the one-channel image D measures the Euclidean CIELAB distance between each pixel and the specimen-specific mean background:

$$D(x, y) = \sqrt{\sum_{c \in \{L^*, a^*, b^*\}} [I_c(x, y) - \bar{b}_c]^2} \tag{2}$$

In Equation (2), $D(x, y)$ is the scalar color difference at pixel (x, y) ; $I_c(x, y)$ is the value in CIELAB channel c ; $\bar{b}_c$ is the corresponding mean background value from Equation (1); and $c \in \{L^*, a^*, b^*\}$ indexes the three CIELAB channels. A larger $D(x, y)$ indicates stronger stain-to-background color separation. D is a pixel-wise color-distance image, not an edge detector or a binary stain mask.

The core relationship is $\tilde{I} \times A = S$: the normalized CIELAB image is multiplied pixel by pixel by the continuous difference weight to form the stain appearance image. Equation (3) states this relationship first, defines the weight A , and then specifies the three channel normalization rules:

$$\begin{aligned}
S_c(x, y) &= \tilde{I}_c(x, y)A(x, y), \quad c \in \{L^*, a^*, b^*\} \\
A(x, y) &= \text{clip}\left(\frac{D(x, y)}{\max(P_{95}(D), \varepsilon)}, 0, 1\right) \\
\tilde{I}_{L^*} &= \frac{I_{L^*}}{100}, \quad \tilde{I}_{a^*} = \frac{I_{a^*}}{128}, \quad \tilde{I}_{b^*} = \frac{I_{b^*}}{128}
\end{aligned} \tag{3}$$

In Equation (3), the first line defines $S_c(x, y)$, channel c of the stain appearance image at pixel (x, y) , as the normalized CIELAB value $\tilde{I}_c(x, y)$ multiplied by weight $A(x, y)$, where c indexes the three CIELAB channels. The second line defines $A(x, y)$ by scaling the difference $D(x, y)$ from Equation (2) with $\max(P_{95}(D), \varepsilon)$ and applying $\text{clip}(\cdot, 0, 1)$, which restricts the weight to $[0, 1]$. Here, $P_{95}(D)$ is the 95th percentile of the same D image, and $\varepsilon = 10^{-6}$ sets a positive lower bound on the denominator to prevent division by zero. Small D values suppress background-like pixels; large values retain more normalized CIELAB information in S . The percentile provides robust scaling; it does not mean that 95% of the image is background.

The third line defines the normalized CIELAB channels using fixed divisors of 100, 128, and 128 for lightness and the two chromatic channels, respectively. Pixel coordinates are omitted in this line for brevity; the same operation is applied at every pixel. The chromatic channels retain their signs: no offset is added, and the three channels are not all mapped to $[0, 1]$. These constants do not depend on image-wise minima, maxima, or statistics fitted to the dataset.

This continuous weighting avoids a manually selected binary threshold and requires no annotated stain contour; however, it does not guarantee that every emphasized pixel is stain-related.

The difference image D is computed from the unscaled CIELAB values in Equation (2) and is supplied to the network without channel normalization. Before injection, the background image from Equation (1) is divided channel-wise by the same factors of 100, 128, and 128. For brevity, this scaled background tensor is also denoted by B in the network description and Equation (4). The network therefore receives S , the unscaled difference D , and the channel-scaled background B . RGB renderings of S and B and false-color renderings of D are used only for presentation and do not create additional model inputs.

For a fixed ROI and border band, let $I(x, y)$ and $\bar{b}$ denote the unscaled three-channel CIELAB pixel and border-mean vectors. Adding the same constant vector Δ to every pixel also shifts the border mean by Δ . Recomputing D then gives

$$D'(x, y) = \| [I(x, y) + \Delta] - [\bar{b} + \Delta] \|_2 = D(x, y)$$

Thus, the entire D image and its 95th percentile remain unchanged. Using the fixed channel scaling in Equation (3),

$$A'(x, y) = A(x, y), \quad S'_c(x, y) = S_c(x, y) + A(x, y) \frac{\Delta_c}{s_c}.$$

Here, primes denote shifted quantities, the subscript 2 denotes the Euclidean norm, and Δ_c is the offset in channel c , with $s_{L^*} = 100$, $s_{a^*} = s_{b^*} = 128$. This cancellation assumes no clipping or quantization. Although D and A remain unchanged, S and B generally change, so classifier invariance does not follow. Background-only errors, localized border contamination, and non-uniform illumination do not satisfy this common-shift condition.

(c) Semantic Meaning and Tensor Dimensions. The three semantic images remain separate because they provide complementary evidence for grading rather than interchangeable views of the same quantity.

B describes the specimen-specific fabric background, D indicates the magnitude of the color difference at each pixel from that background, and S preserves the weighted stain color and spatial pattern. Their dimensions are $B \in \mathbb{R}^{112 \times 112 \times 3}$, $D \in \mathbb{R}^{112 \times 112 \times 1}$, and $S \in \mathbb{R}^{112 \times 112 \times 3}$, respectively.

Before each injection, the selected semantic image is resized to match the height and width of the receiving RGB feature tensor. This alignment changes only spatial resolution and does not change the semantic meaning of B , D , or S ; channel encoding is performed after resizing.

(d) Extraction and Injection Order. The two sequences serve different purposes: extraction follows mathematical dependency, whereas injection controls how the trainable network receives the three cues.

Extraction must follow $B \rightarrow D \rightarrow S$ because B is required to calculate D , and D is subsequently required to construct the weighting map A and stain image S . Network injection follows $S \rightarrow D \rightarrow B$: the trainable pathway first receives local stain appearance, then explicit stain-to-background separation, and finally specimen-specific background context. Table 3 summarizes the definitions and dimensions of the three semantic inputs.

Table 3. Definitions and dimensions of the simplified semantic images.

Semantic Image	Definition	Output Size
S: stain appearance	Normalized CIELAB image weighted by the continuous map A	$112 \times 112 \times 3$
D: color difference	Pixel-wise Euclidean CIELAB distance from the per-image mean background B	$112 \times 112 \times 1$
B: mean background	Outer-border mean CIELAB color repeated over the grid, with channel scaling before injection (Equation (3))	$112 \times 112 \times 3$

3.3. Trainable Feature Pathway and Prediction

The trainable pathway first extracts RGB features, progressively incorporates the three semantic images, and converts the refined feature tensor into nine-grade probabilities. This process consists of three connected steps: (a) lightweight RGB feature extraction, (b) progressive semantic injection in the order $S \rightarrow D \rightarrow B$, and (c) global pooling and nine-grade classification.

(a) Lightweight RGB Feature Extraction Backbone. The trainable RGB branch uses an AlexNet-style early stem [13] to extract low- and mid-level color, contrast, and spatial features. The 11×11 stride-4 convolution and first max-pooling layer rapidly reduce spatial resolution, and the 5×5 convolution expands the channel dimension to 256. Group normalization (GN), rather than batch normalization, follows the trainable convolutional layers because it does not depend on batch-level statistics and is more suitable for small-batch training [14].

The three post-injection refinement stages use depthwise-separable convolution [15]. A depthwise operation filters each channel spatially, and a subsequent 1×1 pointwise operation mixes information across channels. This factorization is substantially smaller than a full convolution with the same input and output widths.

In this manuscript, DSConv and SepConv both denote depthwise-separable convolution: DSConv labels the semantic encoder inside an injection unit, whereas SepConv labels a post-injection refinement layer. Together with the compact channel widths, this design yields exactly 1,403,336 trainable parameters in the complete nine-grade TSI-Net.

(b) Progressive Tri-Semantic Injection. Each semantic image must be aligned with the current RGB feature tensor before fusion. The injection unit encodes the selected semantic image into 32 channels and combines it with the learned RGB feature. Equation (4) presents

the fused output first, defines the supporting semantic feature second, and specifies the fixed injection order in the third line:

$$\begin{aligned} F'_1 &= \text{ReLU}\{\text{GN}[\text{Conv}_{1\times 1}(F_1 \parallel E_1)]\} \\ E_1 &= \text{ReLU}\{\text{GN}[\text{DSConv}_{3\times 3}(\text{Resize}(M_1, H_1, W_1))]\} \\ l &\in \{1, 2, 3\}, \quad (M_1, M_2, M_3) = (S, D, B) \end{aligned} \quad (4)$$

In Equation (4), $l \in \{1, 2, 3\}$ is the injection stage; F_l and F'_l are the RGB features before and after fusion; $M_1 = S$, $M_2 = D$, and $M_3 = B$ specify the fixed injection order; H_l and W_l are the spatial dimensions of F_l ; and E_l is the 32-channel encoded semantic feature. Resize aligns height and width, $\text{DSConv } 3 \times 3$ is a depthwise-separable convolution, $\parallel$ denotes channel-wise concatenation, $\text{Conv } 1 \times 1$ restores the stage width, GN denotes group normalization, and ReLU denotes the rectified linear-unit activation. The first line is the principal fusion relationship; the second defines its supporting semantic feature; and the third explicitly states the fixed injection order. Intermediate fusion allows each cue to influence feature learning rather than only the final classifier.

(c) Global Pooling and Nine-Grade Prediction. After the third refinement stage, final max pooling produces a $5 \times 5 \times 256$ tensor, and global average pooling reduces it to a 256-dimensional vector. Dropout(0.30) is then followed by Dense(9) with Softmax.

The nine probabilities correspond, in order, to grades 1.0, 1.5, 2.0, 2.5, 3.0, 3.5, 4.0, 4.5, and 5.0. At inference, the grade with the highest probability is selected.

3.4. Ordinal Soft-Label Learning

A strict one-hot target treats every incorrect grade as equally wrong, although a neighboring half grade is more similar to the reference than a distant grade [16–18]. Equation (5) therefore converts the expert-assigned reference grade into an ordered Gaussian training-target distribution:

$$q_k^{(y)} = \frac{\exp\left[-\frac{(k-y)^2}{2\sigma^2}\right]}{\sum_{j=1}^K \exp\left[-\frac{(j-y)^2}{2\sigma^2}\right]} \quad (5)$$

In Equation (5), y is the class index of the expert-assigned reference grade; k is a candidate-class index; $q_k^{(y)}$ is the target probability assigned to class k ; j is the summation index used for normalization; $K = 9$ is the number of classes; $\sigma = 0.45$ controls how rapidly the target probability decreases with class distance; and $\exp(\cdot)$ is the exponential function. Class indices 1–9 correspond to Grades 1.0, 1.5, . . . , 5.0, and the denominator makes the 9 target probabilities sum to 1. Equation (5) defines the training target; it is not the final model prediction. Following the ordinal label-smoothing principle adopted by Wang et al. [8] for fabric appearance assessment, the reference grade was assigned the highest target probability, while smaller probabilities were assigned to neighboring grades. Rather than using the fixed 0.70/0.15/0.15 distribution adopted in that study, the present work implemented this principle using a continuous Gaussian target distribution. The smoothing width was fixed at $\sigma = 0.45$ in class-index units. For an interior reference grade, this gives approximately 85.5% target probability to the reference grade and approximately 7.2% to each immediately adjacent grade, with negligible probability assigned to more distant grades. This conservative setting preserves the dominance of the expert-assigned grade while accounting for the perceptual similarity between neighboring half-grade classes. The same value was used in all Gaussian-target experiments. It was fixed before model training and was not selected or adjusted using either the validation set or the test set.

The Gaussian target from Equation (5) is compared with the Softmax output of the model using the cross-entropy loss in Equation (6):

$$L_{CE} = - \sum_{k=1}^K q_k^{(y)} \log p_k \quad (6)$$

In Equation (6), L_{CE} is the training loss; $q_k^{(y)}$ is the Gaussian target probability from Equation (5); p_k is the Softmax-predicted probability for class k ; $K = 9$; and $\log(\cdot)$ denotes the natural logarithm. Minimizing L_{CE} updates the trainable network weights so that the predicted distribution approaches the ordered target distribution. The fixed semantic extractor contains no trainable weights. At inference, the grade associated with the largest p_k is reported.

4. Experimental Setup

4.1. Training Configuration and Comparative Models

The 325 images were randomly divided into 260 training images (80.00%), 32 validation images (9.85%), and 33 test images (10.15%), resulting in an approximately 80%/10%/10% ratio. The training set was used to learn the network parameters, the validation set was used to monitor training and select the model, and the test set was reserved for final performance evaluation. A single random hold-out split was used without cross-validation or an additional stratification procedure. As described in Section 2.1, one source photograph was acquired per physical specimen. No duplicate image was shared among the training, validation, and test sets. The dataset did not include replicate specimens from the same fabric sample or staining/laundrying experiment.

The results reported in Tables 4–6, including those for TSI-Net, were generated using PyTorch 2.10.0 with CUDA 12.8 on a computer equipped with an AMD Ryzen 9 8940HX CPU (2.40 GHz), 16 GB RAM, and an NVIDIA GeForce RTX 5060 GPU (8 GB), running the 64-bit Windows 11 Pro operating system.

Table 4. Accuracy on the common test set and trainable parameter counts for TSI-Net and the baseline convolutional networks.

Method	Acc0 (%)	Acc0.5 (%)	Acc1 (%)	Parameters
TSI-Net	87.88	96.97	100.00	1.4×10^6
AlexNet	72.73	81.82	93.94	5.7×10^7
ResNet18	75.76	81.82	90.91	1.1×10^7
ResNet34	81.82	84.85	93.94	2.1×10^7
InceptionV3	78.79	96.97	100.00	2.2×10^7
VGG16	72.73	84.85	93.94	1.3×10^8

Note: Accuracies were calculated from the reference and predicted grades for the common 33-image held-out test set.

Table 5. Configuration-level comparison of semantic inputs on the common 33-image held-out test set, with trainable parameter counts.

Method	Acc0 (%)	Acc0.5 (%)	Acc1 (%)	Parameters
RGB only	51.52	78.79	90.91	1,007,593
RGB + B	57.58	78.79	90.91	1,082,020
RGB + D	57.58	78.79	87.88	1,168,210
RGB + B + D (without S)	63.64	81.82	90.91	1,242,637
RGB + S + D + B (TSI-Net)	87.88	96.97	100.00	1,403,336

Table 6. Comparison of one-hot and Gaussian soft-label training for TSI-Net on 33 test images.

Method	Acc0 (%)	Acc0.5 (%)	Acc1 (%)
TSI-Net + one-hot training	84.85	93.94	100.00
TSI-Net + Gaussian soft-label learning	87.88	96.97	100.00

All comparative experiments used AdamW [19] with a learning rate of 0.0006, a weight-decay coefficient of 0.0001, and global gradient-norm clipping at 5. The effective batch size was 8, using gradient accumulation over microbatches of at most 4 images. All 6 networks were trained from scratch without pretrained weights, using random seed 42 for a maximum of 40 epochs. No data augmentation was applied to any model.

Tables 4 and 5 report accuracies calculated from the reference and predicted grades for the common 33-image held-out test set. Acc0, Acc0.5, and Acc1 were calculated from the integer numbers of test images satisfying the corresponding absolute-error tolerances and are reported to two decimal places.

AlexNet [13], ResNet18 and ResNet34 [20], InceptionV3 [21], and VGG16 [22] are the comparison networks. Their final classification layers correspond to the nine ordered grades. Model complexity is represented by the parameter counts in Table 4.

All six networks used the same fixed training, validation, and test image lists. RGB ROIs were resized by bilinear interpolation to 224×224 pixels and scaled to [0, 1]. All networks used the same Gaussian target distributions and cross-entropy loss defined in Equations (5) and (6).

Checkpoints were selected according to the lowest validation mean absolute grade error (MAE), with ties resolved first by higher exact-grade accuracy and then by lower hard-label negative log-likelihood. Early stopping used a patience of 10 epochs without improvement under this selection rule. Checkpoint and model selection used only the validation set; the test set was reserved for final evaluation using the metrics in Section 4.2.

The baseline evaluation compared the complete TSI-Net with the five networks listed in Table 4 and compared the five semantic-input configurations summarized in Table 5. Each row in Tables 4 and 5 reports the trainable parameter count and the observed test-set accuracy values for that model or configuration. Section 4.3 describes the configuration-level comparison, whereas Section 4.4 presents the one-hot and Gaussian training-target comparison in Table 6.

4.2. Evaluation Metrics

Because the labels are spaced at 0.5-grade intervals, three tolerance-based accuracies are used. Acc0 counts exact predictions, Acc0.5 counts predictions within one adjacent half-grade, and Acc1 counts predictions within one full grade. Each value is reported as the percentage of evaluated images satisfying the corresponding absolute-error tolerance.

Acc0 is the strictest measure and directly evaluates whether the exact reference grade is reproduced. Acc0.5 and Acc1 indicate whether remaining errors stay within practically neighboring grades. Parameter count measures model size; higher accuracy and fewer parameters indicate better grading performance and a more compact network.

4.3. Semantic-Input Configuration Comparison

Table 5 reports five evaluated semantic-input configurations: RGB only, RGB + B, RGB + D, RGB + B + D (without S), and RGB + S + D + B (TSI-Net). The complete TSI-Net used all three semantic images in the $S \rightarrow D \rightarrow B$ injection order described in Section 3.3.

The Parameters column lists the trainable parameter count for each evaluated configuration—1,007,593, 1,082,020, 1,168,210, 1,242,637, and 1,403,336—in the order

shown in Table 5. Because the configurations have different trainable sizes, Table 5 is a configuration-level comparison rather than a parameter-matched ablation. Accordingly, the difference between RGB + B + D (without S) and the complete TSI-Net reflects both the availability of S and the difference in trainable model capacity; it is not interpreted as an effect attributable only to the fixed S tensor.

4.4. Comparison of Training Targets

The training-target comparison evaluates TSI-Net trained with one-hot targets and with the Gaussian soft targets described in Section 3.4. One-hot targets assign probability 1 to the reference class and 0 to the remaining classes. Gaussian targets distribute probability across the ordered classes according to Equation (5). Cross-entropy in Equation (6) compares each target distribution with the predicted probabilities. Table 6 reports Acc0, Acc0.5, and Acc1 for 33 test images per method, using the definitions in Section 4.2.

5. Results and Discussion

5.1. Comparison with Mainstream Convolutional Networks

Table 4 reports the test set accuracies and model sizes. TSI-Net achieved 87.88% Acc0, 96.97% Acc0.5, and 100.00% Acc1, respectively. Its Acc0 was 6.06 percentage points higher than that of ResNet34 (81.82%). TSI-Net and InceptionV3 achieved the same Acc0.5 and Acc1, and TSI-Net had the smallest parameter count among the compared networks.

5.2. Comparison of Semantic-Input Configurations

Table 5 reports the test-set accuracies of the evaluated semantic-input configurations. RGB only achieved 51.52% Acc0, 78.79% Acc0.5, and 90.91% Acc1. RGB + B and RGB + D each achieved 57.58% Acc0 and 78.79% Acc0.5; their Acc1 values were 90.91% and 87.88%, respectively.

RGB + B + D (without S) achieved Acc0, Acc0.5, and Acc1 values of 63.64%, 81.82%, and 90.91%, respectively. The complete TSI-Net achieved 87.88%, 96.97%, and 100.00%, respectively, which were the highest values among the configurations listed in Table 5.

5.3. Comparison of One-Hot and Gaussian Soft-Label Learning

Table 6 compares TSI-Net trained with one-hot targets and Gaussian soft targets to determine whether encoding grade order is associated with higher observed accuracy under the same training and evaluation protocol.

One-hot training achieved Acc0, Acc0.5, and Acc1 values of 84.85%, 93.94%, and 100.00%, respectively, whereas Gaussian soft-label learning achieved 87.88%, 96.97%, and 100.00%. Thus, the observed Acc0 and Acc0.5 values were each 3.03 percentage points higher with Gaussian soft labels, while Acc1 remained unchanged. These results show higher observed exact-grade and within-half-grade accuracy for Gaussian soft-label learning on this split under the same evaluation protocol.

5.4. Model Complexity Analysis

The fixed CIELAB extractor contains no trainable weights; the semantic injection encoders and fusion layers are trainable. The complete TSI-Net has 1,403,336 trainable parameters, 87.2% fewer than ResNet18's approximately 11.0 million parameters. In Table 5, the complete-model count exceeds the 1,242,637 parameters reported for RGB + B + D (without S) by 160,699. This difference concerns the total trainable sizes of the two configurations and is not a parameter count assigned to the fixed S image. Table 5 therefore compares model configurations together with their respective trainable sizes.

5.5. Discussion, Limitations, and Future Work

On the common held-out test set, TSI-Net achieved the highest Acc0 (87.88%) and shared the highest Acc0.5 (96.97%) and Acc1 (100.00%) with InceptionV3. It also had the lowest parameter count among the compared networks. These results demonstrate favorable exact-grade performance and compactness under the standardized acquisition and evaluation protocol used in this study.

Table 5 reports the observed accuracy ranking together with the parameter count of each configuration. The complete TSI-Net achieved the highest values for all three accuracy metrics and had the largest parameter count within this table. Because the configurations differed in model size, this comparison was not a parameter-matched ablation and did not isolate the contribution of S or evaluate alternative injection orders.

The training-target comparison in Table 6 complements the semantic-image design by examining how the reference grades are represented during learning. Gaussian soft targets assign probability to neighboring classes, whereas one-hot targets concentrate all target probability on the reference class. The Gaussian-trained model achieved higher Acc0 and Acc0.5 under the common evaluation protocol, with unchanged Acc1. Ongoing work is evaluating additional random splits and Gaussian smoothing widths to assess the consistency of this difference and refine the smoothing-width selection.

This study used 325 images acquired under standardized conditions in a professional textile-testing laboratory. The proposed technique is currently being applied within the participating textile research institute, and its evaluation is being extended to industrial partner companies. All specimens in the present dataset were evaluated by one of two professionally trained assessors following the same laboratory grading procedure. Because the assessors evaluated nonoverlapping specimen subsets, the present data do not permit calculation of an inter-rater agreement statistic. Expanded data collection will incorporate dual-assessor evaluation.

The present implementation assumes controlled imaging and a clean outer ROI band that represents the unstained fabric background. Localized border contamination, patterned backgrounds, and non-uniform illumination can alter the semantic representations and are therefore included in the ongoing industrial evaluation. The expanded dataset will cover additional fabric colors, patterns, fiber compositions, soil types, and acquisition environments. Newly collected specimens will be separated from the training and model selection data for independent evaluation, and retraining or fine-tuning will be performed as needed for newly represented conditions. Alternative border widths, background-estimation methods, semantic-injection configurations, Gaussian smoothing widths, and model interpretability analyses will also be investigated. The proposed model is intended to provide a consistent decision support reference for standardized fabric appearance assessment while complementing established laboratory procedures and trained professional judgment.

6. Conclusions

This study developed a lightweight TSI-Net for automatic nine-grade assessment of fabric soil release appearance. From each RGB image, a fixed CIELAB-based procedure constructs the background representation B, color difference representation D, and stain appearance representation S in the order $B \rightarrow D \rightarrow S$, and the trainable backbone incorporates them in the order $S \rightarrow D \rightarrow B$. The complete model contains 1,403,336 trainable parameters. On the common 33-image held-out test set, Gaussian soft-label training achieved Acc0, Acc0.5, and Acc1 values of 87.88%, 96.97%, and 100.00%, respectively, compared with 84.85%, 93.94%, and 100.00% for one-hot training. The observed Acc0 and Acc0.5 values were therefore 3.03 percentage points higher with Gaussian soft labels, while Acc1

remained 100.00% for both target formulations. TSI-Net also achieved the highest Acc0 and shared the highest Acc0.5 and Acc1 among the evaluated baseline networks while using a compact trainable parameter count. Table 5 supports the performance of the complete semantic configuration at the configuration level; because the configurations were not parameter-matched, the comparison is not interpreted as isolating the independent contribution of each semantic input. The proposed technique is currently being applied within the participating textile research institute, and its evaluation is being extended to industrial partner companies. For newly represented fabric backgrounds, patterns, fiber compositions, soil types, and acquisition conditions, representative data will be incorporated into retraining or fine-tuning, and performance will be verified using independent specimens. The model is intended to provide a consistent decision-support reference while complementing established laboratory procedures and trained professional judgment.

Author Contributions: Conceptualization, W.-Y.C., C.-H.H. and L.-W.C.; methodology, W.-Y.C. and C.-H.H.; software, validation, formal analysis, investigation, data curation, visualization, and writing—original draft preparation, W.-Y.C.; resources, C.-H.H.; writing—review and editing, C.-H.H. and L.-W.C.; supervision, C.-H.H. and L.-W.C.; project administration, C.-H.H. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The data used in this study are available from the corresponding author at d1381101@nfu.edu.tw upon reasonable request.

Acknowledgments: The authors thank the professional testing personnel at the Taiwan Textile Research Institute for specimen preparation, imaging, and appearance grading.

Conflicts of Interest: The authors declare no conflicts of interest.

Abbreviations

AATCC, American Association of Textile Chemists and Colorists; Acc0, exact-grade accuracy; Acc0.5, within-half-grade accuracy; Acc1, within-one-grade accuracy; CIELAB, CIE L*a*b* color space; CNN, convolutional neural network; CPU, central processing unit; GAP, global average pooling; ROI, region of interest; TSI-Net, Tri-Semantic Injection Network.

References

1. AATCC TM215-2023; Test Method for Soil Release: Household Soils. American Association of Textile Chemists and Colorists: Research Triangle Park, NC, USA, 2023. Available online: <https://members.aatcc.org/store/tm215/3848/> (accessed on 7 August 2026).
2. Liang, J.; Zhou, J.; Hu, X.; Luo, H.; Cao, G.; Liu, L.; Xiao, K. Digital grading the color fastness to rubbing of fabrics based on spectral reconstruction and BP neural network. *J. Imaging* **2023**, *9*, 251. [CrossRef] [PubMed]
3. Cui, G.; Luo, M.R.; Rhodes, P.A.; Rigg, B.; Dakin, J. Grading textile fastness. Part 1: Using a digital camera system. *Color. Technol.* **2003**, *119*, 212–218. [CrossRef]
4. Cui, G.; Luo, M.R.; Rhodes, P.A.; Rigg, B.; Dakin, J. Grading textile fastness. Part 2: Development of a new staining fastness formula. *Color. Technol.* **2003**, *119*, 219–224. [CrossRef]
5. Luo, M.R.; Cui, G.; Rigg, B.; Dakin, J. A new method for assessing textile fastness based on digital imaging. In *Proceedings of the Second European Conference on Colour in Graphics, Imaging, and Vision (CGIV 2004), Aachen, Germany, 5–8 April 2004*; Society of Imaging Science and Technology: Springfield, VA, USA, 2004; Volume 2, pp. 479–483. [CrossRef]
6. Shao, F.-J.; Xu, P.; Fan, W.; Yan, Y.; Ding, X.; Wu, X. Evaluation of stain release based on image histogram analysis. *Ind. Textila* **2020**, *71*, 204–208. [CrossRef]
7. Peng, Z.; Zhang, C.; Wei, W. Leather defect detection based on improved YOLOv8 model. *Appl. Sci.* **2024**, *14*, 11566. [CrossRef]
8. Wang, J.; Shi, K.; Wang, L.; Li, Z.; Sun, F.; Pan, R.; Gao, W. Automatic assessment of fabric smoothness appearance based on a compact convolutional neural network with label smoothing. *IEEE Access* **2020**, *8*, 26966–26974. [CrossRef]

9. Shi, X.; Cao, W.; Raschka, S. Deep neural networks for rank-consistent ordinal regression based on conditional probabilities. *Pattern Anal. Appl.* **2023**, *26*, 941–955. [CrossRef]
10. Qin, D.; Leichner, C.; Delakis, M.; Fornoni, M.; Luo, S.; Yang, F.; Wang, W.; Banbury, C.; Ye, C.; Akin, B.; et al. MobileNetV4: Universal models for the mobile ecosystem. In *Computer Vision-ECCV 2024*; Leonardis, A., Ricci, E., Roth, S., Russakovsky, O., Sattler, T., Varol, G., Eds.; Lecture Notes in Computer Science; Springer: Cham, Switzerland, 2025; Volume 15098, pp. 78–96. [CrossRef]
11. AATCC M1-2017e; AATCC Standard Reference Detergents and Laundry Detergents in General. American Association of Textile Chemists and Colorists: Research Triangle Park, NC, USA, 2017. Available online: <https://members.aatcc.org/store/m001/2202/> (accessed on 8 September 2026).
12. CIE. *CIE 015:2018 Colorimetry*, 4th ed.; CIE Central Bureau: Vienna, Austria, 2018. [CrossRef]
13. Krizhevsky, A.; Sutskever, I.; Hinton, G.E. ImageNet classification with deep convolutional neural networks. In *Advances in Neural Information Processing Systems 25*; Curran Associates: Red Hook, NY, USA, 2012; pp. 1097–1105.
14. Wu, Y.; He, K. Group normalization. In *Computer Vision-ECCV 2018*; Springer: Cham, Switzerland, 2018; pp. 3–19. [CrossRef]
15. Chollet, F. Xception: Deep learning with depthwise separable convolutions. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 July 2017; pp. 1251–1258. [CrossRef]
16. McCullagh, P. Regression models for ordinal data. *J. R. Stat. Soc. Ser. B* **1980**, *42*, 109–127. [CrossRef]
17. Niu, Z.; Zhou, M.; Wang, L.; Gao, X.; Hua, G. Ordinal regression with multiple output CNN for age estimation. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 27–30 June 2016; pp. 4920–4928. [CrossRef]
18. Cao, W.; Mirjalili, V.; Raschka, S. Rank consistent ordinal regression for neural networks with application to age estimation. *Pattern Recognit. Lett.* **2020**, *140*, 325–331. [CrossRef]
19. Loshchilov, I.; Hutter, F. Decoupled weight decay regularization. In Proceedings of the International Conference on Learning Representations, New Orleans, LA, USA, 6–9 May 2019. Available online: <https://openreview.net/forum?id=Bkg6RiCqY7> (accessed on 7 August 2026).
20. He, K.; Zhang, X.; Ren, S.; Sun, J. Deep residual learning for image recognition. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 27–30 June 2016; pp. 770–778. [CrossRef]
21. Szegedy, C.; Vanhoucke, V.; Ioffe, S.; Shlens, J.; Wojna, Z. Rethinking the Inception architecture for computer vision. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 27–30 June 2016; pp. 2818–2826. [CrossRef]
22. Simonyan, K.; Zisserman, A. Very deep convolutional networks for large-scale image recognition. In Proceedings of the International Conference on Learning Representations, San Diego, CA, USA, 7–9 May 2015. [CrossRef]

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.