**Management and Production Engineering Review**
Volume 17 • Number 1 • March 2026 • pp. 1–17
DOI: 10.24425/mper.2026.158636

# Quality Assessment of Production Using Image Segmentation

Ajay Kumar BOYAT[1], Vinit GUPTA[2]

[1] *Independent Researcher, Indore, India*
[2] *Medicaps University, Department of Electronics Engineering, India*

Received: 05 January 2025
Accepted: 03 November 2025

**Abstract**
Quality assessment of manufactured products is vital to ensure performance, safety, and customer satisfaction across industries. Defects in items such as bottle caps, cables, capsules, leather, and metal components can affect functionality and durability. Traditional inspection methods relying on manual visual checks are time-consuming and error-prone. This study proposes an AI-driven framework using the Probabilistic U-Net integrated with a Conditional Variational Autoencoder (CVAE) for automated defect detection. The model introduces stochastic latent variables to generate multiple plausible segmentation maps, enhancing accuracy under ambiguous or noisy conditions. Using the MVTec Anomaly Detection dataset, which includes defects such as scratches and discoloration, the system applies preprocessing steps including resizing, normalization, and data augmentation to enhance the robustness and consistency of the input data. A hybrid loss combining cross-entropy and Kullback–Leibler divergence improves segmentation precision and latent space alignment. Experimental results confirm robust and reliable defect detection across diverse product categories, demonstrating the model's potential for automated manufacturing quality assurance.

**Keywords**
Conditional Variational Autoencoder, Probabilistic U-Net, Image segmentation, Scratches, Discolouration, and Structural Anomalies.

## Introduction

Quality assessment is essential in manufacturing to ensure that products meet specified standards, protect consumer safety and maintain brand integrity. The basic need for quality assurance arises from various sources, such as reduced production costs, increased customer satisfaction and compliance with regulatory standards (Wu et al., 2020). Increased demand for quality requires strategic, firm quality control that identifies defects and malfunctions before the product reaches the customer (Hsiao et al., 2024). This proactive approach saves money by reducing waste on repeat business and enhancing the company's reputation by providing reliable and safe products to customers (Alzoubi et al., 2022). In recent years, imaging techniques, especially classification, have become essential in quality assessment (Zhao et al., 2020). Segmentation is a tool used to divide an image into discrete parts to better identify features or defects in a product (Peng et al., 2019). This technique provides manufacturers with a comprehensive view of product features for accurately detecting and classifying defects, such as cracks, scratches, or irregular patterns that may degrade quality and may be impossible. Using advanced imaging techniques, developers can automate quality assurance processes, improve efficiency, accuracy and significantly reducing the risk of human error (Islam et al., 2024).

Image processing techniques can identify various kinds of damage to merchandise throughout production or earlier than attaining the purchaser. These encompass surface cracks, corrosion, misalignments, floor discolouration, and irregular edges (Crognale et al., 2023). Computer imaginative and prescient algorithms discover invisible cracks, while colourimetric analyses spotlight corrosion in metals and fabrics. Spatial dating algorithms flag nonconforming components, whilst spectral imaging detects floor discolouration, indicating high-quality troubles. Edge-detection algorithms become aware of irregular edges to enhance safety and satisfaction (Tariq et al., 2021). These multidimensional analysis strategies can enhance satisfactory evaluation protocols across diverse production sectors (Cheon et al., 2021).

***Corresponding author**: Ajay Kumar Boyat – Independent Researcher, Indore, India, Postal Address House 74, Palhar Nagar, Opposite Garden 1, 60 Feet Road, Airport Road, Indore, M.P., India. 452005, phone: (+91) 9926612382, e-mail: drajaykumarboyat@gmail.com*

Deep getting-to-know fashions have modified the photograph processing game for quality assessment tasks like photo segmentation. The models that have attracted the most attention for their superb overall performance in semantic segmentation tasks are U-Net (Rajamani et al., 2023) and Conditional Variational Autoencoders (CVAEs) (Celard et al., 2023).

The U-Net structure, to begin with, is designed for biomedical picture segmentation; it has a symmetrical growing route for precise localization and a contracting course for acquiring context. The community can compile a comprehensive picture of capabilities at multiple stages because this encoder-decoder design enables the renovation of inherent spatial hierarchies throughout the segmentation technique (Zhang et al., 2021). Incorporating bypass connections in U-Net complements the transfer of high-resolution traits from the encoder to the decoder, for this reason, augmenting first-class detail visibility in the segmented output (Zioulis et al., 2022). Conditional Variational Autoencoders (CVAEs) enhance the capability of conventional autoencoders by conditioning the latent variable modelling on supplementary enter, which is mainly helpful in eventualities with more than one practicable segmentation due to photo ambiguity.

The advent of Probabilistic U-Nets marks a widespread advancement in segmentation methodologies, especially for ambiguous or noisy datasets. Unlike conventional U-Nets, which aim to yield a deterministic segmentation output, Probabilistic U-Nets incorporate stochastic factors into the model that permit them to expect multiple possible segmentations for a single center (Kohl et al., 2018). This is accomplished through a combination of latent variable sampling and generative modelling principles, in which the version learns to express uncertainty in its predictions. The key difference lies in how those models deal with uncertainty and variability when entering data. Probabilistic U-Nets now offer a factored estimate and deliver a distribution over viable segmentations, reflecting inherent ambiguities present in visual records. This characteristic allows operators to evaluate the reliability of segmentation consequences, an essential component in best evaluation scenarios wherein the fee of misclassification may be significant.

These fashions can improve defect detection frameworks in practical packages by incorporating learned uncertainty estimates into decision-making tactics. For example, regions identified as ambiguous with the aid of the version can be prioritized for similar inspection, optimizing helpful resource allocation. Additionally, probabilistic approaches enable persistent mastering of new statistics, improving version adaptability and accuracy over the years and fostering continuous development in producing excellent standards (Kohl et al., 2018; Carpanzano and Knüttel 2022; Hossain et al., 2024).

## Literature review

In (Mirra et al., 2020) the method used to detect and identify abnormalities in fruits to reduce health risks was fruit analysis using images. This process includes preprocessing, segmentation, feature extraction, and classification based on colour, texture, size, appearance, and defect tests. Provides a complete description. A method in (Dharmik et al., 2022) proposed developing a computer model that uses digital image processing and machine learning techniques to monitor grain quality. The approach included visualisation and machine learning techniques, extracting features from grain seeds, and developing a neural network model to enhance the results.

According to Basnet (2017), a machine learning–based approach was used to evaluate how accurately model predictions align with human assessments of the quality of printed models. The subjective test and the automated objective model were part of the analysis. According to the study, the most reliable indicator of a product's quality is the contrast between the printed pattern and its outer edge. For the detection quality, the final model achieved an accuracy of 83%, performing beyond the range of the RMS measurement. The study by Schmitt et al. (2020) investigates an approach to quality control in industrial manufacturing using predictive models and edge cloud computing. All four steps – data collection, implementation, sampling, and implementation – are part of the scoping process. The comprehensive strategy includes all four steps – data collection, processing, modelling, and deployment. The results demonstrate that this strategy can significantly decrease inspection volumes, leading to economic benefits and an improvement in the supply of high-quality, defect-free products. As demonstrated in (Schömig-Markiefka et al., 2021), artefacts can significantly affect the precision of deep learning models used for prostate cancer detection. In particular, the presence of artefact-contaminated regions in histology slides introduces noise and variability that adversely influence feature extraction and classification performance. This sensitivity to artefacts highlights the importance of robust preprocessing and quality control mechanisms to ensure reliable and accurate histopathological analysis. It discovered that model performance could be drastically affected by any severity-dependent artefact. The study suggests strategies to prevent model accuracy losses and stress-testing using synthetically generated artefacts for clinical validation.

Medeiros et al. (2021) investigates the potential of convolutional neural network (CNN) based deep learning models for X-ray image-based crambe seed quality

monitoring. When classifying seeds according to germination, vigour, and interior tissue integrity, the models attained 91.5%, 95.5%, and 82% accuracy, respectively. This suggests that digital radiographic images can provide valuable information on crambe seeds. A method in Mishra and Passos (2021) presents a strategy for improving spectral data processing deep learning models using chemometrics expertise. It recommends utilizing spectral data augmentation and Hotelling's T2 and Q statistics for pre-filtering outliers. The method was evaluated on a real-world NIR dataset to forecast mango dry matter. The data demonstrated enhanced predictive performance with an RMSEP of just 0.79%. Ding et al. (2021) identifies the best image processing algorithm by comparing eleven full-reference IQA models. Four basic vision tasks – denoising, deblurring, super-resolution, and compression – are used to train the models. By doing subjective tests on the improved images, we can see how well each model performs perceptually and identify their strengths and weaknesses. The research calls for improved benchmarks and suggests ways to lessen the likelihood of overfitting in future iterations of quality assessment models.

## Preliminaries

Let $x$ represent the ground-truth segmentation map of interest, and y represent the production images as observed input. In contrast to the standard U-Net design, the Probabilistic U-Net goes beyond it by introducing a latent space to model segmentation uncertainties. The Loss function of Variational Autoencoder (VAE) loss function, which can be written as:

$$L = \mathbb{E}_{z\sim q_\Phi(z|x)}\left[\log p_\theta\left(x|z,y\right)\right] KL(q_\Phi\left(z|y\right) \parallel p\left(z\right)) \quad (1)$$

A variational latent variable $z$ is modelled using a reparameterizable distribution function $q_\Phi(z|y)$, where $\Phi$ stands for the inference network parameter. Based on the observed input $y$ and the sampled latent variable $z$. The segmentation output $x$ is conditioned by the generative model. The segmentation map $p_\theta(x|z,y)$. To get close to the actual posterior $p_\theta(x|y)$ the framework uses the objective function provided to optimize a variational lower limit.

This is the Evidence Lower Bound (ELBO) used in Variational Autoencoders (VAEs) – a deep generative model that learns to represent data (such as images, sounds, etc.) in a compressed latent space $z$. The goal is to maximize this objective (or equivalently minimize the negative ELBO), which balances reconstruction accuracy and latent regularization.

$$q_\Phi(z|y) - \text{Encoder/Inference Model} \quad (2)$$

This is an approximate posterior that models how likely latent variable $z$ is, given some observed input $y$. Parameterized by a neural network with parameters $\Phi$. Acts as the encoder that maps input data to a latent representation.

$$p_\theta(x|z,y) - \text{Decoder/Generative Model} \quad (3)$$

This is the likelihood or reconstruction term, data $x$ can be generated given the latent variable $z$ and possibly conditional input $y$. Parameterized by a neural network with parameters $\theta$. It represents the decoder reconstructing $x$ from $z$.

$p(z)$ Prior Distribution over latent variables assumed to be a standard normal distribution $\mathcal{N}(0,1)$. Acts as a regularizes ensuring that the latent space is smooth and well-behaved.

The expected log-likelihood or reconstruction loss $\mathbb{E}_{z\sim q_\Phi(z|x)}\left[\log p_\theta\left(x|z,y\right)\right]$. Measures how well the reconstructed data $\hat{x}$ matches the original data $x$. Encourages the decoder to generate realistic data from the latent code $z$.

The Kullback–Leibler divergence term.

$$KL\left(q_\Phi(z|y) \parallel p(z)\right) \quad (4)$$

Measures how much learned distribution $q_\Phi(z|y)$ deviates from the prior $p(z)$. Acts as a regularize term pushing the encoded latent variables to be close to the prior (e.g., Gaussian). Intuitively keep the latent space organized and prevent overfitting. Intuitive Meaning the overall objective combines two goals:

$$L = \text{Reconstruction Accuracy} - \text{Regularization Cost} \quad (5)$$

The first term encourages the model to reconstruct input data accurately. The second term penalizes large deviations of the latent space from a prior distribution. By optimizing $L$, the model learns both a compressed latent representation $z$. A generative model capable of producing new samples similar to the training data.

### Connection to Conditional VAE (CVAE)

The conditional term $y$ in $p_\theta(x|z,y)$ and $q_\Phi(z|y)$, this is a Conditional VAE, where generation is conditioned on some label or attribute $y$ (including attributes such as class labels and style). This allows the model to generate outputs conditioned on specific properties. This allows the model to generate outputs conditioned on specific properties.

During inference, the model generates a distribution of segmentation maps by sampling multiple latent variables z, capturing aleatoric uncertainty caused by variability in the input data. This enables probabilistic confidence estimates for each pixel in the segmentation map, making the framework particularly effective for decision-making in uncertain or variable production environments. The Probabilistic U-Net provides a flexible and efficient approach to uncertainty-aware image segmentation by implicitly modelling the posterior without requiring a closed-form solution.

## Conditional Variational Autoencoder

A rudimentary knowledge of variational autoencoders (VAEs) is necessary to comprehend the proposed technique. VAEs are crucial to clarifying the complexities of the suggested method. Combining the concepts of autoencoders and variational inference, variational autoencoders (VAEs) are robust, unsupervised generating models. The objective is to gather a representation of complex high-dimensional input data into a low-dimensional latent space. The latent area encapsulates the essential shape and variability of the statistics as a non-stop multivariate distribution. The two main additives of Variational Autoencoders (VAEs) are the encoder and the decoder. The decoder reconstructs records from the latent area to the authentic input area, even as the encoder maps the statistics entered into the latent area.

In variational autoencoders (VAEs), input data is transformed into probability distributions concerning the latent variables in preference to being encoded without delay as a singular factor in the latent area. More flexibility and uncertainty modelling are made possible by this probabilistic representation. With the help of the decoder network, VAEs may recreate the input statistics while also generating new samples based on the learnt opportunity distributions in the latent representation.

The basic idea behind VAEs is to estimate the distribution (i.e. marginal probability) of the input data as represented by $p_\theta(x)$. To this end, VAEs increase the level of evidence. ELBO has two parts: the regularization step, which adjusts the latent space distribution to match a prior distribution (usually a multivariate Gaussian distribution), and the reconstruction loss, which measures the VAE's ability to reconstruct input data to achieve a delicate balancing act accurately.

$$\log P_\theta(X) \geq L(\theta, \Phi; X) = \mathbb{E}_{q_\Phi(z|x)}[-\log q_\Phi(z|x) + \log P_\theta(x, z)] \quad (6)$$

Equation (6) represents the Evidence Lower Bound (ELBO) formulation in Variational Inference (VI), which is the mathematical foundation of Variational Autoencoders (VAEs).

Equation (6) shows that the log marginal likelihood (or log evidence) of data $X$ – i.e., $\log P_\theta(X)$ – is intractable to compute directly. Instead, we compute a lower bound on it, denoted $L(\theta, \Phi; X)$, which we can maximize instead. This lower bound is known as the Evidence Lower Bound (ELBO). So, the Equation (6) states:

$$\log P_\theta(X) \geq \text{ELBO} = L(\theta, \Phi; X) \quad (7)$$

Maximizing the ELBO helps us indirectly maximize the likelihood of the observed data. Whereas $X$ Observed image. $z$ is latent variable states the hidden variable that explains how $X$ is generated. $P_\theta(x, z)$ is Joint probability of data and latent variable which represent Approximate posterior. $\theta$ is model parameters (decoder) such as Parameters of the generative model.

$\Phi$ is model parameters (encoder) Parameters of the inference (approximation) model.

$$L(\theta, \Phi; X) = \mathbb{E}_{q_\Phi(z|x)}[\log P_\theta(x, z) - \log q_\Phi(z|x)] \quad (8)$$

where, $L(\theta, \Phi; X)$ is lower bound on log-likelihood.

Now, substitute the joint probability term $P_\theta(x, z) = P_\theta(x|z) P(z)$. So, the Equation (8) becomes:

$$L(\theta, \Phi; X) = \mathbb{E}_{q_\Phi(z|x)}[\log P_\theta(x|z)] - KL(q_\Phi(z|x) \parallel P(z)) \quad (9)$$

This is the same form as the Variational Autoencoder loss derived earlier. The true log-likelihood of the data:

$$\log P_\theta(x) = \log \int P_\theta(x|z) P(z) \mathrm{d}z \quad (10)$$

Equation (10) states true data likelihood (intractable) $\log P_\theta(x)$ is hard to compute because integrating over all possible $z$ is intractable. So, we introduce a tractable approximate distribution $q_\Phi(z|x)$ and rewrite:

$$\log P_\theta(x) = \log \int q_\Phi(z|x) \frac{P_\theta(x, z)}{q_\Phi(z|x)} \mathrm{d}z \quad (11)$$

Applying Jensen's inequality gives the lower bound: $\log P_\theta(x) \geq \mathbb{E}_{q_\Phi(z|x)} \left[\log \frac{P_\theta(x, z)}{q_\Phi(z|x)}\right]$

That expectation is precisely the ELBO:

$$L(\theta, \Phi; X) = \mathbb{E}_{q_\Phi(z|x)} [\log P_\theta(x, z) - \log q_\Phi(z|x)] \quad (12)$$

The ELBO represents two competing goals:

1. Reconstruction accuracy – how well the model can explain observed data given latent variables.

$$\mathbb{E}_{q_\Phi(z|x)}[\log P_\theta(x|z)]$$

2. Regularization – how close the approximate posterior is to the prior distribution.

$$-KL(q_\Phi(z|x) \parallel P(z))$$

Maximizing the ELBO encourages the encoder–decoder pair to:

- Learn a meaningful latent representation $z$.
- Generate realistic samples $x$ from the latent code.

VAEs transforms an intractable Bayesian inference problem into an optimizable lower bound that can be computed via stochastic gradient descent using neural networks.

Reformulating Equation (1) with the ELBO loss on the left side produces the following result. The Kullback–Leibler divergence, denoting the reconstruction and regularization components, is presented on the right side of the equation.

To describe the observed data distribution using latent variables in an unsupervised manner, the conditional variational autoencoder (CVAE) (Ivanov et al., 2019) was used, on which an encoder and decoder are built. The CVAE, as illustrated in Fig. 1. When the distribution of $p_\theta(x|y)$ is multimodal. It can perform better than the deterministic model. A deterministic regression model with Mean Square Error (MSE) will forecast the average blurry value for $x$, if $x$ is real. Various realistic items can be sampled from the $x$ distribution examined using CVAE (Kingma et al., 2014). As with VAE, conditioning all distributions taken into consideration at $x$ yields the lower bound of variational for CVAE as follows:

$$L_{\text{CVAE}}(x, y; \theta, \Psi, \Phi) = \mathbb{E}_{q_\Phi(z|x,y)}[\log p_\theta(x|z, y)] - D_{\text{KL}}(q_\Phi(z|x, y) \parallel p_\Psi(z|y)) \leq \log p_{\theta,\Psi}(x|y) \quad (13)$$

Equation (13) is a conditional form of the Variational Autoencoder (VAE) Evidence Lower Bound (ELBO), commonly referred to as the Conditional Variational Autoencoder (CVAE) loss. The CVAE is an extension of the Variational Autoencoder (VAE) that incorporates conditional information $y$ (e.g., class labels, attributes, or context) into both the encoder and decoder networks. This allows the model to generate or reconstruct data conditioned on specific attributes – an example, generating an image of a "dog" conditioned on the label "dog".

| Symbol | Meaning | Description |
|---|---|---|
| $x$ | Observed data | Input data (e.g., image, sentence) |
| $y$ | Conditional variable | Label, attribute, or context |
| $z$ | Latent variable | Hidden or encoded representation |
| $D_{\text{KL}}(\cdot \parallel \cdot)$ | KL divergence | Regularization term measuring divergence between distributions |

**ELBO Definition for CVAE**

$$L_{\text{CVAE}}(x, y; \theta, \Psi, \Phi) = \mathbb{E}_{q_\Phi(z|x,y)}[\log p_\theta(x|z, y)] - D_{\text{KL}}(q_\Phi(z|x, y) \parallel p_\Psi(z|y)) \quad (14)$$

This is the conditional evidence lower bound on the log-likelihood $\log p_{\theta,\Psi}(x|y)$:

$$L_{\text{CVAE}}(x, y; \theta, \Psi, \Phi) \leq \log p_{\theta,\Psi}(x|y) \quad (15)$$

1. Reconstruction Term:
   $\mathbb{E}_{q_\Phi(z|x,y)}[\log p_\theta(x|z, y)]$
   - Encourages the decoder to reconstruct $x$ accurately from the latent variable $z$ and conditional input $y$.
   - This is equivalent to minimizing reconstruction error (e.g., cross-entropy or MSE).

   Intuition: Given latent code $z$ and condition $y$, can the model generate $x$ correctly.
2. KL Divergence Term:
   $D_{\text{KL}}(q_\Phi(z|x, y) \parallel p_\Psi(z|y))$
   - Regularizes the latent space by forcing the posterior $q_\Phi(z|x, y)$ to be close to the conditional prior $p_\Psi(z|y)$.
   - Keeps the latent distribution smooth and prevents overfitting.

   Intuition: Make sure the learned latent representation doesn't deviate too far from a prior distribution conditioned on $y$.
3. Upper Bound:
   $L_{\text{CVAE}}(x, y; \theta, \Psi, \Phi) \leq \log p_{\theta,\Psi}(x|y)$
   - The ELBO (left-hand side) lower bounds the true log-likelihood of the data conditioned on $y$.
   - Maximizing the ELBO approximates maximizing the actual conditional log-likelihood – since direct computation of $\log p_{\theta,\Psi}(x|y)$ is intractable.

The CVAE learns two things simultaneously:

First, an encoder (inference model): $q_\Phi(z|x,y)$ – maps observed data and condition to a latent variable.

Second, a decoder (generative model): $p_\theta(x|z,y)$ – reconstructs or generates data given $z$ and condition $y$.

The conditional prior $p_\Psi(z|y)$ acts as a flexible prior distribution that can adapt based on the condition.

In practice:

- $q_\Phi(z|x,y)$ and $p_\Psi(z|y)$ are modelled as Gaussian distributions with mean and variance predicted by neural networks.
  $z = \mu_\Phi(x,y) + \sigma_\Phi(x,y) \odot \epsilon, \epsilon \sim \mathcal{N}(0,I)$.

Loss function minimized during training (negative ELBO):

$$\begin{aligned}\mathcal{L}_{\text{loss}} = & -\mathbb{E}_{q_\Phi(z|x,y)}[\log p_\theta(x|z,y)] \\ & + D_{\text{KL}}(q_\Phi(z|x,y) \parallel p_\Psi(z|y))\end{aligned} \tag{16}$$

This equation formalizes how a Conditional Variational Autoencoder (CVAE) learns to:

- Generate data $x$ conditioned on $y$,
- While maintaining a structured latent space $z$,
- By maximizing the conditional evidence lower bound (ELBO).

Utilizing the parameterization technique to optimize CVAE aims. Remember that the neural network with parameter $\Psi$ represents the prior distribution $p\Psi(z|y)$, which depends on $y$. Whereas VAE only uses two trainable neural networks, CVAE uses three. We will propose CVAE modifications in our study, including hybrid models with relevant functions obtained from resampling regression and Gaussian stochastic neural networks.

Unsupervised learning-oriented CVAEs strive to construct significant representations of incoming data by modelling the fundamental probability distributions. Probabilistic U-Net enables the execution of tasks such as segmentation within a variational framework, extending the variational capabilities of Variational Autoencoders (VAEs) to supervised learning.

## Materials & Methods

Image segmentation is extracting specific gadgets or capabilities from an image. Among the most famous methods of example segmentation is the U-NET model, which has many potential packages past its unique area of organic image segmentation, including Earth sciences and area studies. Feeding an image right into a deep convolutional neural network (CNN) named U-NET will generate a segmentation map as its output to help in picture translation. In supervised gaining knowledge, the model is taught to map incoming photos to their corresponding segmentations, which requires providing corresponding segmentation masks. The deterministic nature of U-NET is a downside despite its incredible performance in picture segmentation responsibilities. The entire loss of attention for resources of uncertainty and stochasticity in the deterministic mapping from input pics to output segmentation maps can lead to overfitting and bad generalization to clean records.

A new Probabilistic U-Net version for photograph segmentation was proposed with the aid of Kohl et al. (2018) and integrates a Conditional Variational Auto-Encoder (CVAE) with the U-NET (Kingma et al., 2014; Sohn et al., 2015). Figure 1 shows the architecture of the suggested model. In particular, the samples taken from the VAE's latent feature space are used to condition the U-NET's segmentations. During the evaluation phase, "what-if" scenarios can be assessed within this two-dimensional plane, which encompasses all possible variations in segmentation. The model may produce many segmentation maps from a single input image by basing segmentation generation on the latent space. Each map represents a distinct area of the sampled latent feature space. According to the authors, the model may learn hypotheses with a low probability and predict them with the corresponding frequency (Hossain et al., 2024).

The corresponding segmentation is generated by merging and transmitting the green U-NET output block and the blue z-sampled latent space block to the red F block. In this context, $S_i$ represents the segmentation that coincides with the $z_i$ latent space sample, and $\theta$ and $\Phi$ denote the U-NET and F model parameters, respectively. As part of the training process, the model is designed to achieve two objectives: first, to accurately partition wildfires using the input data, and second, to generalize well to scenarios that are uncommon or unexpected. The model is constrained to achieve the first objective by decreasing the supervised cross-entropy loss between the produced segmentation, $S(X,z)$, and the actual ground truth, $Y$. The model gains more information by decreasing the KL divergence between the posterior $Q(z|X,Y)$ and prior $P(z|X)$ distributions of the variables in the latent feature space. Total loss is, hence, the outcome of combining the two losses.

$$\begin{aligned}L(X) = & \mathbb{E}_{Q_\Phi(Z|X)}[-\log P_\Psi(Y|X,Z)] \\ & + \beta KL[Q_\Phi(Z|X) \parallel P(Z|X)]\end{aligned} \tag{17}$$

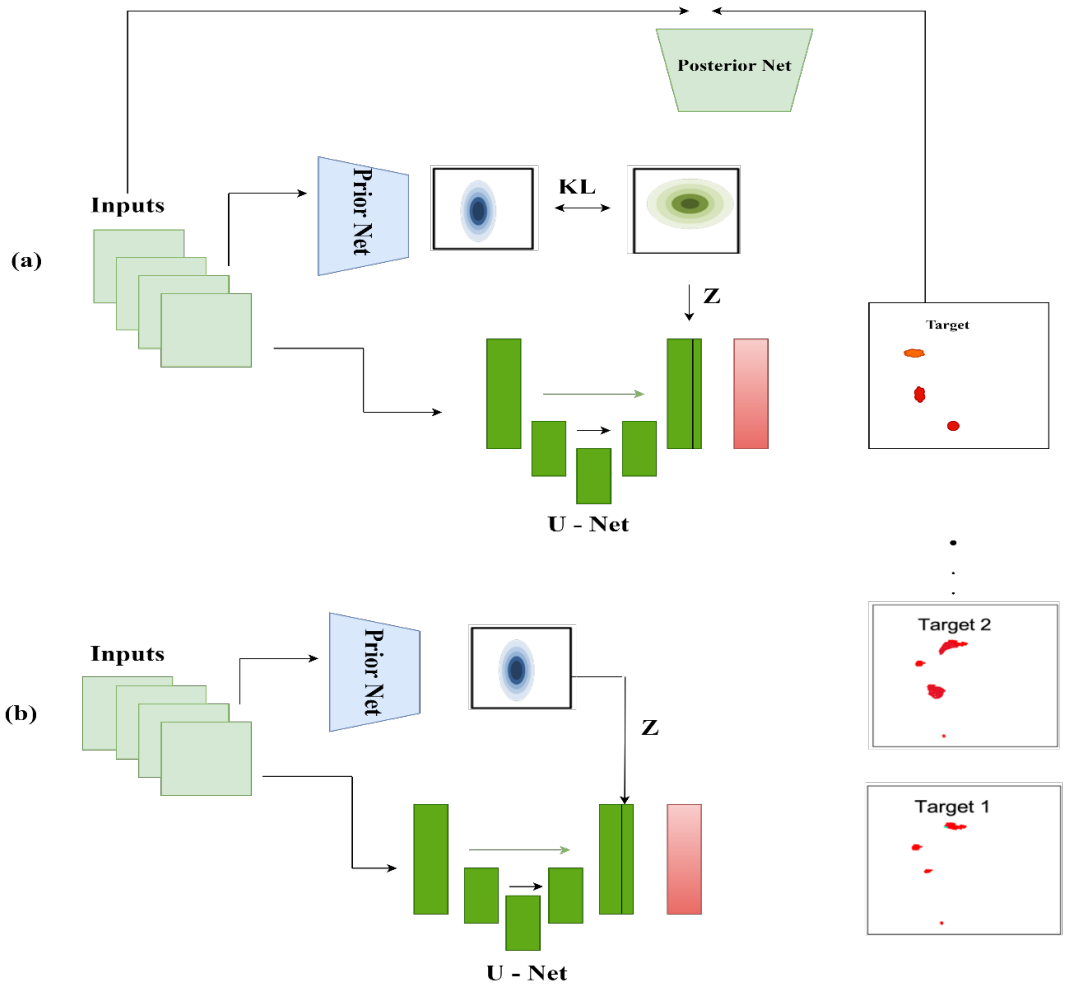



Fig. 1. The suggested Probabilistic U-Net architecture is depicted graphically. (a) shows the training strategy in which outputs are encoded into multivariate Gaussian distributions by the posterior network and inputs into the prior network. The U-Net outputs are joined with samples from the unified multivariate Gaussian distribution to generate target data stochastic events. (b) shows the inference technique that uses the previous network to gather samples.

Equation (17) used in Variational Autoencoders (VAEs), but this specific formulation seems to include an additional variable (like $Y$ or $S$), suggesting it's part of a Conditional Variational or Structured Variational model often used in supervised or semi-supervised settings.

This equation defines a variational loss function that balances two terms:

1. Reconstruction or prediction error: how well the model can predict or generate $Y$ given $X$ and latent variable $Z$.
2. Regularization (KL divergence): how close the approximate posterior $Q_\Phi(Z|X)$ is to a prior $P(Z|X)$.

The coefficient $\beta$ (beta) controls the trade-off between the two – this is called a $\beta$-VAE when applied in unsupervised contexts.

| Symbol | Meaning | Description |
|---|---|---|
| $X$ | Input data | Example: image, sentence, or feature vector |
| $Y$ | Output label or target variable | Example: class label or predicted property |
| $Z$ | Latent variable | Encoded representation learned from data |
| $KL[\cdot \parallel \cdot]$ | Kullback–Leibler divergence | Measures difference between two distributions |
| $\beta$ | Regularization weight | Adjusts influence of the KL term |

**Reconstruction / Prediction Loss**

$\mathbb{E}_{Q_\Phi(Z|X)}[-\log P_\Psi(Y|X,Z)]$

Represents the expected negative log-likelihood of predicting $Y$ given $X$ and latent variable $Z$. Encourages the decoder (parameterized by $\Psi$) to accurately predict or reconstruct the output. Equivalent to minimizing cross-entropy or mean-squared error, depending on the model.

Intuition: The model should predict $Y$ correctly from the latent code $Z$ and input $X$.

**Regularization (KL Divergence) Term**

$\beta KL[Q_\Phi(Z|X) \parallel P(Z|X)]$

Measures how different the learned latent distribution $Q_\Phi(Z|X)$ is from the prior $P(Z|X)$. Keeps the latent space organized, smooth, and prevents overfitting. The $\beta$ coefficient controls how strongly we regularize:

$\beta = 1$: standard VAE loss.

$\beta > 1$: encourages disentangled latent features ($\beta$-VAE).

$\beta < 1$: allows more flexibility for reconstruction.

Intuition: Ensures latent variables stay close to a prior (like Gaussian) for stability and generalization. The total loss combines two objectives:

$$L(X) = \underbrace{\text{Prediction Error}}_{\text{Reconstruction term}} + \beta \times \underbrace{\text{Latent Regulation}}_{\text{KL divergence}}$$

The first term ensures good predictions. The second term ensures that the learned latent representations $Z$ remain well-structured and meaningful.

In summary, this equation describes a variational learning objective that aims to:

- Learn meaningful latent variables $Z$,
- Predict $Y$ accurately from $X$ and $Z$,
- Maintain a smooth, regularized latent space through the KL divergence.

It's a generalized variational framework used in conditional generative models and information bottleneck architectures.

The model's output is controlled by the hyperparameter $\beta$, which determines the extent to which the KL-divergence component, also called the regularization term, contributes. The model is trained from beginning to end. Logarithmic scaling was used to optimize the hyperparameter from $10^{-6}$ to $10^{-3}$, and 0.0001 is the ideal value for $\beta$.

## Data Collection and Preprocessing

This study uses the MVTec Anomaly Detection data set, widely recognized as robust in the quality analysis industry. The dataset contains images of 15 products, such as bottles, fibers, and metal nuts, as well as corresponding descriptions for error-prone samples. The detailed descriptions of this dataset make it a capable source that relies on it for training classification models aimed at quality assessment of the materials.

The MVTec dataset contains images that simulate real-world faults in manufacturing, including scratches, scratches, discoloration, and missing features These variables are important for developing potential defect-related models of many kinds in a variety of unseen things. Key features of the data set include:

1. High-quality images: Ensures high-quality imperfections.
2. Ground truth masks: Provide pixel-level annotations for fault locations, which are important for supervised training of the segmentation model.
3. Different classes: The mix of different elements brings variation, which helps make the model more interdisciplinary.

Flaws: The statistics set covers subtle and vital flaws, permitting the model to stumble on subtle symbols. To streamline the education procedure, the dataset turned into partitioned into education, validation, and trying out subsets. Approximately 70% of the images were allotted for training, 15% for validation, and 15% for checking out. This ensures a comprehensive assessment of the model's efficacy on each visible and unseen information.

For the cause of schooling and comparing fashions, this look at hired preprocessing strategies to trade the format of the uncooked statistics. To assure consistency and compliance with the U-Net architecture, the initial snap shots have been contracted to a hard and fast resolution of 256×256 pixels. To make certain the model trained quickly and the gradient updates have been strong, the depth values of the pixels have been normalized to the interval [0, 1]. To make the model more immune to overfitting and enhance its common robustness, facts augmentation tactics had been used. This blanketed rotation, flipping, scaling, and brightness changes.

Mask training worried resizing floor truth masks to suit enter pics' dimensions and changing them into binary formats. Train-check splitting ensured a balanced distribution of defects and everyday samples, stopping records leakage. Oversampling strategies have been applied to the faulty class to address the imbalance between ordinary and faulty samples, making sure enough exposure to defects all through education while improving the version's capability to pick out outliers in practical settings (Fig. 2).

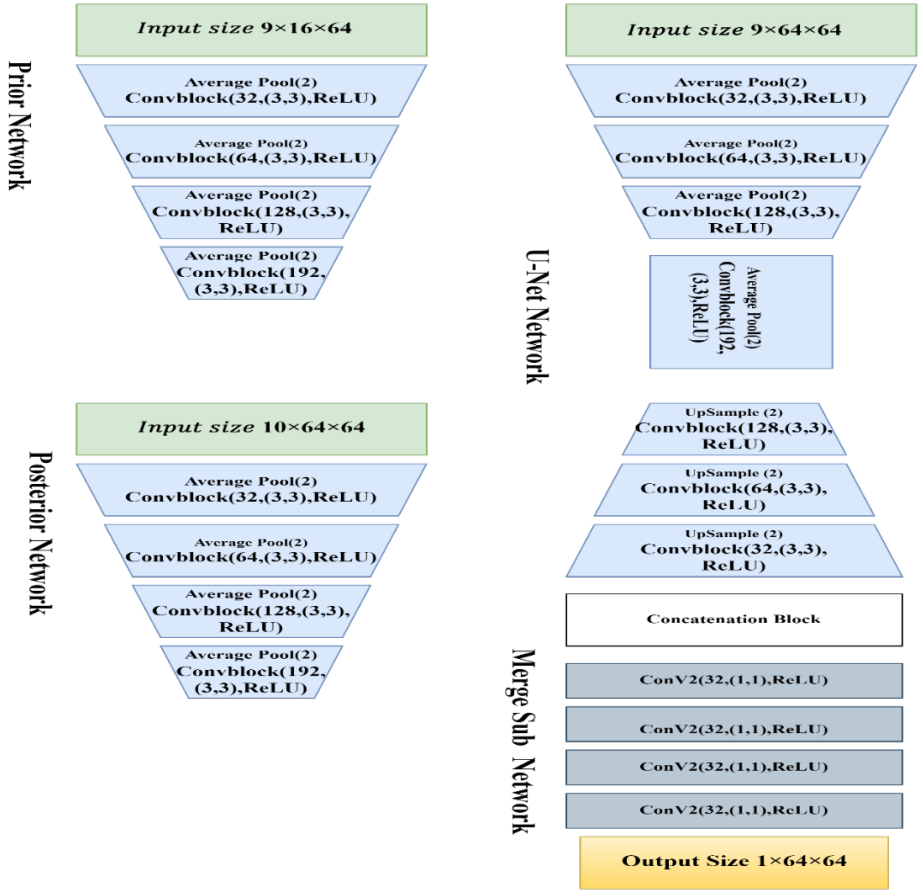



Fig. 2. Three primary networks comprise the Probabilistic U-Net architecture: U-Net (with a merging sub-network), the previous network, and the posterior network.

## Model architecture

An ideal framework for quality assessment using image classification is a probabilistic U-Net framework. It adds a conditional variation auto-coder (CVAE) to the deterministic U-Net model, which generates several possible segmentation maps for the input image considering the uncertainty during segmentation U-net is the main segmentation backbone using the classic encoder, decoder, architecture implements. CVAE includes a hidden feature space, which contributes to stochasticity in the segmentation process. The input image is placed in a latent Gaussian distribution by the first grid. This classification reflects the heterogeneity of the input data and is used to generate different segmentation maps. By inputting the encoding image and the corresponding ground truth segmentation, the background network resolves the hidden space. This allows the model to search for hidden signals that better match the actual classification system.

The posterior network's latent space is utilized to draw samples during training, while the prior network is used to sample latent variables z during inference. One distinct theory regarding the segmentation is reflected in each sample. Prior to passing the combined features to the U-Net decoder, the latent variable z is joined with the feature maps produced by the U-Net encoder. By basing the segmentation on the sampled latent variable, this approach enables the decoder to produce numerous plausible segmentation maps that correspond to various portions of the latent space that were sampled. Supervised Cross-Entropy Loss and Kullback–Leibler Divergence are used in tandem to train the model end-to-end. The former reduces the gap between the ground truth masks and predicted segmentation maps, while the latter aligns the posterior and prior distributions in latent space. This hybrid framework ensures high segmentation accuracy and generalization capability, making it particularly useful for quality assessment in scenarios where defect boundaries are ambiguous or stochastic.

**Algorithm 1. Probabilistic U-Net with CVAE for Image Segmentation**

**Require:** Input image $X$, ground truth segmentation $Y$, learning rate $\eta$, regularization parameter $\beta$, number of epochs $E$

**Ensure:** Segmentation output $S(X, z)$

1. **Initialize:** Model parameters $\theta$(U-Net) and $\Psi$(CVAE), prior network $P(z|X)$, and posterior network $Q(z|X, Y)$.
2. **for** epoch $= 1$ to $E$ **do**

   *Step 1:* U-Net Feature Extraction

   Encode input image $X$ using the U-Net encoder to obtain feature maps $F_{\text{enc}}(X)$.

   *Step 2:* Latent Space Construction

   Compute the prior distribution $P(z|X)$ from $F_{\text{enc}}(X)$.

   Compute the posterior distribution $Q(z|X, Y)$ using $F_{\text{enc}}(X)$ and $Y$.

   *Step 3:* Sampling from Latent Space

   During training, sample latent variable $z \sim Q(z|X, Y)$.

   During inference, sample $z \sim P(z|X)$.

   *Step 4:* Conditional Decoding

   Concatenate $F_{\text{enc}}(X)$ and $z$.

   Decode the concatenated features using the U-Net decoder to generate the segmentation map $S(X, z)$.

   *Step 5:* Loss Computation

   Compute the supervised cross-entropy loss: $L_{\text{CE}} = -\sum_i Y_i \log(S(X, z)_i)$

   Compute the KL divergence: $L_{\text{KL}} = KL(Q(z|X, Y) \parallel P(z|X))$

   Combine total loss: $L_{\text{total}} = L_{\text{CE}} + \beta L_{\text{KL}}$

   *Step 6:* Parameter Update

   Update parameters $\theta$ and $\Psi$ using gradient descent

   $\theta, \Psi \leftarrow \theta, \Psi - \eta \nabla_{\theta,\Psi} L_{\text{total}}$
3. **end for**
4. **Output:**

   Final segmentation map $S(X, z)$ for each input image $X$.

## Results and Discussion

The integration of stochastic U-Net model through Conditional Variational Autoencoders allows the model to generate multiple plausible segmentation outcomes, which enhances robustness against ambiguous or noisy input data. This capability is crucial for real-world manufacturing environments where defect appearances can vary widely. Furthermore, the probabilistic outputs provide uncertainty estimates that can guide decision-making, resource allocation, and reduce misclassification costs. To describe various tables that illustrate the conditions and states of different objects and materials – bottle caps, toothbrushes, cables, capsules, leather, metal nuts, transistors, and zippers. Each table details specific damages, alterations, or defects that these objects can undergo, alongside a description of their "Original Image" state.

Table 1 demonstrates the different bottle caps in their states. In the "Original Image", the cap is complete, with no damages at all. The "Broken Small" image depicts a small cracked or chipped cap. The "Contamination" image contains the cap exposed to foreign materials and dirt that can easily make the cap dirty. Finally, the "Broken Large" image depicts a bottle cap that is highly damaged, such as large cracks, dents, or missing pieces, which might make it unusable. Each image is a vivid representation of what a bottle cap might go through.

Table 1
Categories of Product Quality Defects OF Bottle cap

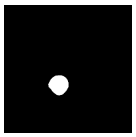
| | |
|---|---|
| Original Image |  |
| Broken Small |  |
| Contamination |  |
| Broken Large |  |

Table 2 compares two conditions of a toothbrush and is categorized as follows: "Original Image": The "Original Image" shows a toothbrush in its faultless, brand-new state, with no visible defects and fully functional design. "Defective": The "Defective" image presents a toothbrush that has sustained some form of damage or flaw, such as bent bristles, broken parts, or other imperfections that affect its usability or appearance. This table is helpful in illustrating the distinction between a healthy, completely intact toothbrush and one that

is no longer so. This table displays a number of cable states by the type of fault or alteration. The "Original Image" shows how the cable was designed and as built initially. The "Cable Swap" image depicts one which has been swapped for, or otherwise altered from its normal condition, perhaps representing another model or type. In the "Combined" the cable is likely to result from several faults that co-occurred. The "Inner Cut Insulation" photo features a cable with damaged inner insulation, exposing its internal wires. The "Outer Cut Insulation" photo features a cable whose external protective layer is damaged. The "Missing Cable" photo represents a partly absent cable. The "Missing Wire" photo features a cable missing one or more wires, making it partially useless. The "Poke Insulation" picture is of a cable with insulation punctured or damaged, thereby possibly exposing the inner wiring. Lastly, the "Bent Wire" picture shows a cable whose wire has been physically bent or deformed, possibly impairing its conductivity. This table quite effectively visualizes several possible problems that cables can have.

Table 2
Categories of Product Quality Defects of Tooth brush

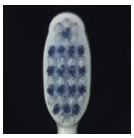
| Item | Original Image | Defective Image |
|---|---|---|
| Tooth Brush |  |  |

Table 3 displays a number of cable states by the type of fault or alteration. The "Original Image" shows how the cable was designed and as built initially. The "Cable Swap" image depicts one which has been swapped for, or otherwise altered from its normal condition, perhaps representing another model or type. In the "Combined" the cable is likely to result from several faults that co-occurred. The "Inner Cut Insulation" photo features a cable with damaged inner insulation, exposing its internal wires. The "Outer Cut Insulation" photo features a cable whose external protective layer is damaged. The "Missing Cable" photo represents a partly absent cable. The "Missing Wire" photo features a cable missing one or more wires, making it partially useless. The "Poke Insulation" picture is of a cable with insulation punctured or damaged, thereby possibly exposing the inner wiring. Lastly, the "Bent Wire" picture shows a cable whose wire has been physically bent or deformed, possibly impairing its conductivity. This table quite effectively visualizes several possible problems that cables can have.

Table 3
Categories of Product Quality Defects of Cable

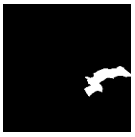
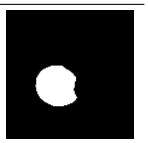
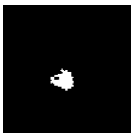
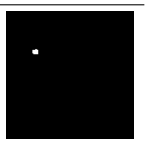
| | |
|---|---|
| Original Image |  |
| Cable Swap |  |
| Combined |  |
| Cut Inner Insulation |  |
| Cut Outer Insulation |  |
| Missing Cable |  |
| Missing Wire |  |
| Poke Insulation |  |
| Bent Wire |  |

Table 4 shows extraordinary states of a capsule, labelled in line with numerous types of damage or flaws. The "Original Image" represents the pill as first meant, without harm or marks. The "Fault Imprint" image contains an indentation or imprint that might imply a few kinds of stress or effect that has disturbed the floor of the pill. The "Poke" picture shows a hollow in a capsule punctured by using sharp objects or external forces. The "Scratch" image displays clean surface abrasions or scratches that may have been inflicted

at some stage in friction or via difficulty dealing with. The "Squeeze" image describes a flattened or deformed capsule wherein compression turned into a demonstration that the shape had lost some of its integrity. The ultimate picture is that of a "Crack". The capsule in this image is cracked, which could jeopardize its use or may be protection. This table will show the numerous steps or forms of harm a pill may preserve.

Table 5 illustrates several conditions of leather, classified according to different types of damage or alteration. The "Original Image" is the image of the leather in its original, undamaged state, as it was designed. The "Cut" image is leather that has been cut or torn, leaving an obvious cut or tear in the material. The "Fold" picture displays leather that has been flexed or wrinkled, meaning the image is going to bear noticeable folds which might influence the overall look or feel. The "Glue" picture contains images of leather that has been bonded with glue, which will also reflect marks of adhesiveness that affect the outer layer. The "Poke" picture features leather images which bear small punctures or holes caused by pointed objects or impact. Finally, the "Color" is that where the leathers have suffered a color change either from fading and staining, or intentional dyeing. This table shows the ways and means of how leather changes or deteriorates over time.

Table 6 illustrates various conditions of a metal nut, classified via exceptional forms of alterations or damage. The "Original Image" shows the nut in its perfect, undamaged kingdom, because it changed into initially manufactured. The "Color" image suggests a trade within the nut's coloration, which will be because of oxidation, staining, or intentional coating. The "Flip" photo represents a nut that has been flipped or circled, probably indicating a shift in its orientation or alignment. The "Scratch" picture illustrates visible surface abrasions or scratches on the metallic, usually resulting from friction or outside touch. Lastly, the "Bent" photo highlights a steel nut that has been deformed, with substantive bends that can affect its functionality. This table effectively provides distinctive ranges or forms of damage that a steel nut might also revel in over time.

Table 4
Categories of Product Quality Defects of Capsule

| | |
|---|---|
| Original Image |  |
| Fault Imprint |  |
| Poke |  |
| Scratch |  |
| Squeeze |  |
| Crack |  |

Table 5
Categories of Product Quality Defects of Leather

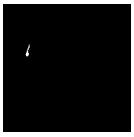
| | |
|---|---|
| Original Image |  |
| Cut |  |
| Fold |  |
| Glue |  |
| Poke |  |
| Color |  |

Table 6
Categories of Product Quality Defects of Metal nut

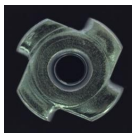
| | |
|---|---|
| Original Image |  |
| Color |  |
| Flip |  |
| Scratch |  |
| Bent |  |

Table 7 represents exceptional situations of a transistor which can be categorized underneath various kinds of damage or defects. The "Original Image" is that of the transistor in its excellent, un-broken form, as manufactured. The "Cut Lead" picture is of a transistor with a lead wire cut or severed, possibly affecting its capability. The "Damaged Case" image depicts a transistor where its outer casing is visibly damaged and cracked, dented, or otherwise compromised. The "Misplaced" image is that of a transistor with a lead or component which is misaligned and potentially causing improper connections and, therefore, performance. Finally, the "Bent Lead" image is an image of a transistor in which one or more lead wires are bent or distorted, which can affect the electrical connection or stability of that transistor. This table quite well represents the various damage or problems that a transistor may experience.

Table 8 shows several conditions of a zipper, grouped by the type of damage. The "Original Image" is the zipper in its original, undamaged, fully functional state with no visible damage. The "Combined" image probably represents a zipper with several problems occurring together, such as damaged teeth and a compromised slider. The "Fabric Border" image reveals that the fabric lining the border of the zipper is torn and may therefore impair the stability of the zipper. The "Fabric Interior" image indicates that the lining of the inner fabric lining around the zipper is torn or worn and will affect its overall performance. The "Rough" image shows that the teeth or mechanism of the zipper may have become rough due to a wearing or friction effect thus failing to open or shut properly. The "Split Teeth" image shows some misaligned or separated teeth in the zipper that is preventing the zipper from proper meshing. The image "Squeezed Teeth" shows that this type of teeth has an effect of compression or deformation thus preventing smooth closing in the zipper. Lastly, the "Broke Teeth" image shows a zipper with broken or missing teeth, thus rendering the zipper not to function properly. From this table, it becomes evident how a zipper may be damaged or compromised.

The LIDC-IDRI dataset's GED metric trend is displayed in Figure 3 and Table 9. Although the proposed ACP-U-Net variants show marginal differences in GED values (ranging from $0.218 \pm 0.220$ to $0.242 \pm 0.168$). The observed improvement, while numerically small, may still contribute to practical benefits in real-world defect detection scenarios. Specifically, a lower GED value indicates a more consistent and precise segmentation of defect regions, which can enhance operator confidence in automated decision-making processes. However, it should be noted that the differences across model variants are relatively minor and may not necessarily translate into statistically significant improvements in overall detection accuracy.

Table 7
Categories of Product Quality Defects of Transistor

| | |
|---|---|
| Original Image |  |
| Cut Lead |  |
| Damaged Case |  |
| Misplaced |  |
| Bent Lead |  |

Table 8
Categories of Product Quality Defects of Zipper

| | |
|---|---|
| Original Image |  |
| Combined |  |
| Fabric Border |  |
| Fabric Interior |  |
| Rough |  |
| Split Teeth |  |
| Squeezed Teeth |  |
| Broke Teeth |  |

Table 9
Comparison of Model Variants Based on Generalized Energy Distance (GED) Metrics

| Model Variant | GED (mean ± std-dev) |
|---|---|
| ACP-U-Net (AA) | $0.242 \pm 0.168$ |
| ACP-U-Net (FC) | $0.220 \pm 0.188$ |
| ACP-U-Net (Mixture AA) | $0.238 \pm 0.191$ |
| ACP-U-Net (Mixture FC) | $0.218 \pm 0.220$ |

Statistical Validation and Discussion of Performance Differences: To assess whether the observed differences in Generalized Energy Distance (GED) among ACP-U-Net variants represent meaningful improvements or are within the margin of experimental variability, a series of pairwise statistical tests were conducted. Specifically, two-tailed paired t-tests ($\alpha = 0.05$) were applied to compare the GED values across identical test samples for all model pairs.

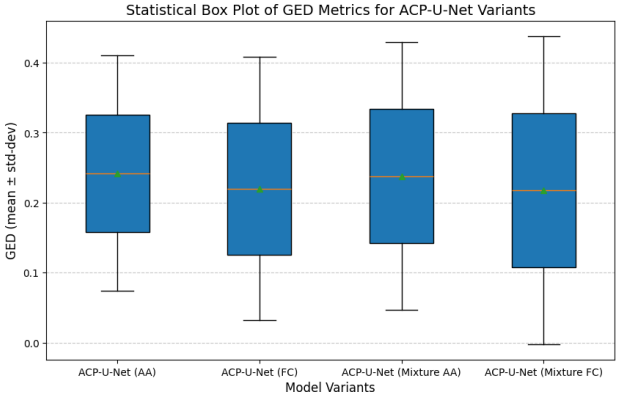



Fig. 3. Statistical analysis of the GED metrics for ACP-U-Net Variants.

The results showed that none of the pairwise comparisons achieved statistical significance (all $p > 0.05$), indicating that the observed variations in GED – although numerically small (ranging from $0.218 \pm 0.220$ to $0.242 \pm 0.168$) – fall within the range of experimental variability. Thus, while the ACP-U-Net (Mixture FC) achieved the lowest mean GED, the difference cannot be interpreted as a statistically confirmed performance improvement over other variant.

These findings suggest that the architectural modifications (Attention Aggregation vs. Feature Concatenation and their Mixtures) do not yield substantial differences in segmentation consistency. In practical terms, the slightly lower GED values correspond to marginal gains in the precision of defect boundary segmentation, but these improvements are unlikely to produce significant changes in real-world defect detection accuracy or decision-making confidence.

Interpretation and conclusion of results: While the proposed probabilistic ACP-U-Net architecture demonstrates reliable performance and stable uncertainty modelling, the statistical analysis indicates that the differences among its variants are not statistically significant. Therefore, the claims regarding superior performance over baseline approaches have been tempered to reflect the limited evidence for strong differentiation among model variants.

Future work will incorporate larger datasets, cross-validation across multiple domains, and additional statistical measures (e.g., Wilcoxon signed-rank tests, ANOVA) to confirm whether these small improvements are reproducible and practically meaningful in

industrial defect detection applications. To better assess the robustness of these findings, future work will include statistical significance testing (e.g., paired t-tests or Wilcoxon signed-rank tests) across multiple datasets and experimental runs. This will help determine whether the observed variations in GED represent meaningful improvements rather than random fluctuations due to model stochasticity or data variability.

When looking at the GED measure, the ACP-U-Net with the Mixture FC variation stands out as the top performer. In every case, models using the Mixture FC version outperformed the gold standard Probabilistic U-Net (AA). Figure 3 illustrates that Mixture FC variant in the ACP-U-Net architecture provides an optimal trade-off between overlap and diversity, making it the top-performing model for the GED metric. The prediction of ACP-U-Net with Mixture FC was closest to the reference segmentation distributions with higher overlap and diversity that better approximate inter-observer variability. Furthermore, although mixtures-based models typically achieved stronger overlap, they were slightly less divergent than their more tractable counterparts. This cost is illustrated graphically in Appendix B. The improved modeling ability of ACP-U-Net with Mixture FC makes it a reasonable candidate for challenging medical imaging segmentation tasks.

However, a limitation of the generalized ACP-U-Net framework, including Mixture FC, is the increased number of hyperparameters that require tuning for optimal performance. As an illustration, in mixed models, finding the ideal number of components (N) and the temperature ($\tau$) for the Gumbel-Softmax distribution, which regulates the relaxation of the discrete categorical distribution, is essential. In this scenario, optimizing the model for a certain dataset is a very time-consuming process. One possible line of future work is learning the number of mixture components instead of fixing it at the outset of training, thus not requiring manual tuning and flexibility.

## Conclusions

This research advances automated manufacturing quality assessment by combining detailed defect categorization with state-of-the-art probabilistic segmentation models. The extensive cataloguing of defects across multiple product types provides a valuable reference for understanding common failure modes and their visual characteristics. The ACP-U-Net with Mixture FC variant emerges as a leading model, offering enhanced segmentation accuracy and uncertainty quantification, thereby addressing limitations of deterministic methods. The implementation of the Probabilistic U-Net model, particularly the ACP-U-Net variant with Mixture Fully Connected (Mixture FC) layers, demonstrates superior performance in defect segmentation tasks. Table 1–Table 9 and Figure 3 show that this model variant achieves the best GED metrics on the LIDC-IDRI dataset, indicating an optimal balance between segmentation overlap and diversity. This suggests that the Mixture FC model effectively captures inter-observer variability and uncertainty inherent in defect detection.

While the ACP-U-Net (Mixture FC) achieved the lowest GED ($0.218 \pm 0.220$), suggesting slightly better segmentation consistency, the overall performance differences among variants were modest. Therefore, these results should be interpreted as indicative of incremental rather than substantial gains. Future research will incorporate formal statistical analyses to confirm whether such improvements are significant and to better understand their impact on real-world defect detection reliability and confidence in decision-making systems.

The probabilistic framework's ability to generate diverse segmentation hypotheses enables more reliable detection of defects, even under challenging conditions. This leads to improved quality assurance, reduced misclassification, and better allocation of inspection resources. The approach sets a new benchmark for automated defect detection, with broad applicability across manufacturing sectors. The detailed defect categorizations across various product types underscore the importance of tailored inspection criteria and highlight the challenges in standardizing quality control processes. The study's findings advocate for the adoption of AI-driven, probabilistic segmentation frameworks to improve inspection reliability and efficiency.

This study introduces a new method for evaluating manufacturing quality using the Probabilistic U-Net framework, which combines the strengths of U-Net and Conditional Variational Autoencoders. By integrating stochastic elements into the segmentation process, the model addresses the limitations of deterministic architectures, such as poor generalization and susceptibility to overfitting. The framework's ability to generate multiple plausible segmentation maps enhances defect detection, even in ambiguous and noisy scenarios. Utilizing the MVTec Anomaly Detection dataset, the study demonstrates the model's robustness in identifying diverse product defects. The incorporation of uncertainty estimates into segmentation outputs ensures reliability in decision-making, thereby improving resource allocation and minimizing misclassification costs.

Future work should focus on optimizing latent space representations and developing adaptive learning mechanisms to enhance model scalability and generalizabil-

ity. Additionally, expanding the dataset diversity and integrating real-time inspection systems could further improve practical deployment. Establishing standardized evaluation protocols based on this framework will facilitate industry-wide adoption and continuous improvement in product quality monitoring. This probabilistic framework has the ability to impact many different industries since it establishes a new standard for automated quality assessment in production. Future work could explore further optimization of the latent space and adaptive learning mechanisms to enhance scalability and adaptability across broader use cases.

**Acknowledgments**

The authors sincerely acknowledge the editors and reviewers for their insightful comments, constructive feedback, and timely responses, which significantly improved the quality of this manuscript.

## References

Alzoubi H.M., Ahmed G., & Alshurideh M. (2022). An empirical investigation into the impact of product quality dimensions on improving the order-winners and customer satisfaction. *International Journal of Productivity and Quality Management*, *36*(2), 169–186. DOI: 10.1504/IJPQM.2022.124711.

Basnet R. (2017). Automated Quality Assessment of Printed Objects Using Subjective & Objective Methods Based on Imaging and Machine Learning Techniques. *M.S. thesis, Rochester Inst. of Technology, Rochester, NY, USA*.

Carpanzano E. & Knüttel D. (2022). Advances in Artificial Intelligence Methods Applications in Industrial Control Systems: Towards Cognitive Self-Optimizing Manufacturing Systems. *Applied Sciences*, *12*(21). Article 10962. DOI: 10.3390/app122110962.

Celard P., Iglesias E.L., Sorribes-Fdez J.M., Romero R., Vieira A.S., & Borrajo L. (2023). A survey on deep learning applied to medical images: from simple artificial neural networks to generative models, *Neural Computing and Applications*, *35*(3), 2291–2323. DOI: 10.1007/s00521-022-07953-4.

Cheon M., Yoon S.J., Kang B., & Lee J. (2021). Perceptual image quality assessment with transformers. *Proceedings of the IEEE/CVF Computer Society Conference on Computer Vision and Pattern Recognition Workshops*, 433–442. Open Access with No DOI found.

Crognale M., De Iuliis M., Rinaldi C., & Gattulli V. (2023). Damage detection with image processing: a comparative study. *Earthquake Engineering and Engineering Vibration*, *22*(2), 333–345. DOI: 10.1007/s11803-023-2172-1.

Dharmik R.C., Chavhan S., Gotarka S.r, & Pasoriya A. (2022). Rice quality analysis using image processing and machine learning. *3C TIC. Cuadernos de Desarrollo Aplicados a las TIC*, *11*(2), 158–164, 2022.

Ding K., Ma K., Wang S., & Simoncelli E.P. (2021). Comparison of Full-Reference Image Quality Models for Optimization of Image Processing Systems, *International Journal of Computer Vision*, *129*(1),1258–1281.

Hossain M.I., Amin M.Z., Anyimadu D. T., & Suleiman T.A. (2024). Comparative Study of Probabilistic Atlas and Deep Learning Approaches for Automatic Brain Tissue Segmentation from MRI Using N4 Bias Field Correction and Anisotropic Diffusion Pre-processing Techniques. *arXiv preprint arXiv:2411.05456*. DOI: 10.48550/arXiv.2411.05456.

Hsiao L., Ma X., & Chen Y-J. (2024). The Effects of Selling Formats and Upstream Competition on Product Pricing and Quality Design. *Manufacturing and Service Operations Management*, *26*(1), 1526–1541. DOI: 10.1287/msom.2022.0470

Islam M.R., Zamil M.Z.H., Rayed M.E., Kabir M.M., Mridha M.F., Nishimura S., & Shin J. (2024). Deep Learning and Computer Vision Techniques for Enhanced Quality Control in Manufacturing Processes. *IEEE Access*, *12*, 121449–121479. DOI: 10.1109/ACCESS.2024.3453664.

Ivanov O., Figurnov M., & Vetrov D. (2019). Variational autoencoder with arbitrary conditioning. *International Conference on Learning Representations (ICLR 2019)*, 1–25.

Kingma D.P., Rezende D.J., Mohamed S., & Welling M. (2014). Semi-supervised learning with deep generative models. *Advances in Neural Information Processing Systems (NeurIPS)*, *4*(January), 3581–3589.

Kohl S.A.A., Romera-Paredes B., Meyer C., Fauw J.D., Ledsam J.R., Maier-Hein K.H., Eslami S.M.A., Rezende D.J., &. Ronneberger O. (2018). A probabilistic U-net for segmentation of ambiguous images. *Advances in Neural Information Processing Systems*, NeurIPS, 6965–6975.

Medeiros A.D.D., Bernardes R.C., Silva L.J.D., Freitas B.A.L.D., Dias D. C.F.D.S., & Silva C.B.D. (2021). Deep learning-based approach using X-ray images for classifying Crambe abyssinica seed quality. *Industrial Crops and Products*, *164*, 113378.

Mirra K.B., Pooja P., Ranchani S., & Kumari R.R. (2020). Fruit Quality Analysis using Image Processing. *International Journal of Engineering and Advanced Technology (IJEAT)*, *9*(5), 88–91.

Mishra P. & Passos D. (2021). Chemometrics and deep learning improved the predictive performance of near-infrared spectroscopy models for dry matter prediction in mango fruit, *Chemometrics and Intelligent Laboratory Systems*, *212*, 104287.

Peng Y., Ruan S., Cao G., Huang S., Kwok N., & Zhou S. (2019). Automated Product Boundary Defect Detection Based on Image Moment Feature Anomaly. *IEEE Access*, *7*, 52731–52742. DOI: 10.1109/ACCESS.2019.2911358

Rajamani K.T., Rani P., Siebert H., ElagiriRamalingam R., & Heinrich M.P. (2023). Attention-augmented U-Net (AA-U-Net) for semantic segmentation. *Signal, Image Video Processing*, *17*(1), 981–989. DOI: 10.1007/s11760-022-02302-3.

Schmitt J., Bönig J., Borggräfe T., Beitinger G., & Deuse J. (2020). Predictive model-based quality inspection using Machine Learning and Edge Cloud Computing. *Advanced Engineering Informatics*, *45*, 101101.

Schömig-Markiefka B., Pryalukhin A., Hulla W., Surname, I., Bychkov A., Fukuoka J., Madabhushi A., Achter V., Nieroda L., Büttner R., Quaas A., & Tolkach Y. (2021). Quality control stress test for deep learning-based diagnostic model in digital pathology. *Modern Pathology*, *34*(12), 2098–2108.

Sohn K., Yan X., & Lee H. (2015). Learning structured output representation using deep conditional generative models, *Advances in Neural Information Processing Systems (NeurIPS)*, *28* (January), 3483–3491.

Tariq N., Hamzah R.A., Ng T.F., Wang S.L., & Ibrahim H. (2021). Quality Assessment Methods to Evaluate the Performance of Edge Detection Algorithms for Digital Image: A Systematic Literature Review. *IEEE Access*, *9*, 87763–87776. DOI: 10.1109/ACCESS.2021.3089210.

Wu J., Xing B., Si H., Dou J., Wang J, Zhu Y., & Liu, X. (2020). Product Design Award Prediction Modeling: Design Visual Aesthetic Quality Assessment via DCNNs. *IEEE Access*, *8*, 211028–211047. DOI: 10.1109/ACCESS.2020.3039715.

Zhang Z., Jiang T., Liu C., & Zhang L. (2021). An Effective Classification Method for Hyperspectral Image With Very High Resolution Based on Encoder–Decoder Architecture, *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, *14*, 1509–1519. DOI: 10.1109/JSTARS.2020.3046245.

Zhao Y., Hao K., He H., Tang X., & Wei B. (2020). A visual long-short-term memory based integrated CNN model for fabric defect image classification. *Neurocomputing*, *380*, 259–270. DOI: 10.1016/j.neucom.2019.10.067.

Zioulis N., Albanis G., Drakoulis P., Alvarez F., Zarpalas D., & Daras P. (2022). Hybrid Skip: A Biologically Inspired Skip Connection for the UNet Architecture. *IEEE Access*, *10*, 53928–53939. DOI: 10.1109/ACCESS.2022.3175864.