Triad: Empowering LMM-based Anomaly Detection with Expert-guided Region-of-Interest Tokenizer and Manufacturing Process

Supplementary Material

A. Extra Results

A.1. Results on PCB-Bank

We also test results on PCB-Bank [33], which includes a bunch of PCB images in a real industrial scene. Because part of the images in PCB-Bank come from VISA [41] and RealIAD [30], which are overlapped with our training data, we use the rest subset of PCB-Bank to evaluate for maintaining the zero-shot setting. Detecting defects on PCB is complicated and needs a lot of experience and knowledge in electronics, which is lacking in the general LMMs' knowledge base. However, Triad still outperforms due to its ability to use extra knowledge from the detailed manufacturing processes of PCB products.

Table A1. Zero-shot Anomaly Detection Performance with **manufacturing process** on PCB-BANK [33].

Model	Params	PCB-Bank (Subset)	
		0-shot	+ MFG Proc.
Qwen2-VL [31]	2B	49.9%	61.0% (11.1%↑)
LLava-1.6 [22]	7B	38.9%	59.9% (21.0%↑)
MiniCPM-V [34]	8B	61.2%	61.0% (0.2%↓)
LLaVA-OneVision-si [16]	8B	60.8%	61.0% (0.2%↑)
LLaVA-OneVision-ov [16]	8B	59.6%	60.9% (1.1%↑)
Qwen2-VL [31]	7B	59.5%	61.1% (1.6%↑)
Qwen2-VL [31]	72B	59.6%	61.0% (1.4%↑)
LLaVA-OneVision-ov [16]	72B	60.9%	61.4% (0.5%↑)
Myriad [19]	7B	61.8%	61.8% (0.0%↑)
Triad-llava-1.6	7B	63.6%	64.8% (1.2%↑)
Triad-ov	7B	62.8%	63.8% (1.0%↑)

A.2. Results on MMAD

We report results on MMAD (**categories for training are filtered out**) in Tab. A2. Triad still achieves the best performance on MMAD and outperforms LLaVA-OV by 9% in both 0/1-shot. Note that the references for 1-shot evaluation on MMAD (especially for GoodsAD) are not actually the same product as the query image. Therefore, all models do not improve much under 1-shot setting.

A.3. Extensive qualitative results

In this section, more qualitative examples are provided to further illustrate how the manufacturing process helps anomaly detection and how Triad interacts with the changing manufacturing processes.

Table A2. 0/1-shot anomaly detection performance on MMAD.

Model	Qwen2-VL-7B	Qwen2.5-VL-7B	LLaVA-OV-ov	Triad-ov
0-shot	56.25%	50.14%	60.74%	**70.22%**
+ MFG Proc.	49.50%↓	45.58%↓	59.60%↓	**71.02%**↑
1-shot	57.78%	42.71%	60.80%	**70.92%**

In Fig. A1, we provide an example of the cable in MVTec-AD. The full manufacturing process is shown in Tab. A8. Four responses from Triad are presented: one includes the full process, and three omit certain information. The steps involving copper wire drawing and insulation extrusion do not specify the number of wires; therefore, their omission does not affect Triad's judgment regarding the missing wire cable. However, when the cable assembly step indicates there are three cables and explains how they are assembled, Triad adjusts its detection and considers the visible black-colored hole (the absence of a wire) as part of the design.

Another example from PCB-Bank [33] is present in Fig. A2. The models are tested in two settings: zero-shot anomaly detection (green box) and anomaly detection with manufacturing process information (pink box). In the zero-shot anomaly detection setting, all models fail to correctly identify the missing button, although Triad detects a missing substrate that might be part of the design. This indicates a need for anomaly detection that incorporates information from the manufacturing process. The anomaly detection with manufacturing process information is presented in the pink boxes. In this example, Triad demonstrates its superior ability to utilize information from the PCB manufacturing process and precisely identify the missing button.

A.4. Failure case study

Figure A4 presents four failure cases revealing Triad's current limitations: (a) **Viewpoint ambiguity:** Requires 3D scans/X-ray imaging for better detection; (b) **Design-related errors:** Need CAD/GERBER design files for resolution; (c) **Label-threshold conflict:** Defects below a prescribed size are labeled as normal by the dataset; (d) **Texture blindness:** Complex patterns obscure flaws in textiles, which necessitates texture-aware training data.

B. More details for evaluation

Evaluation prompts are following templates shown in Tab. A6. For general LMMs and our method, questions and

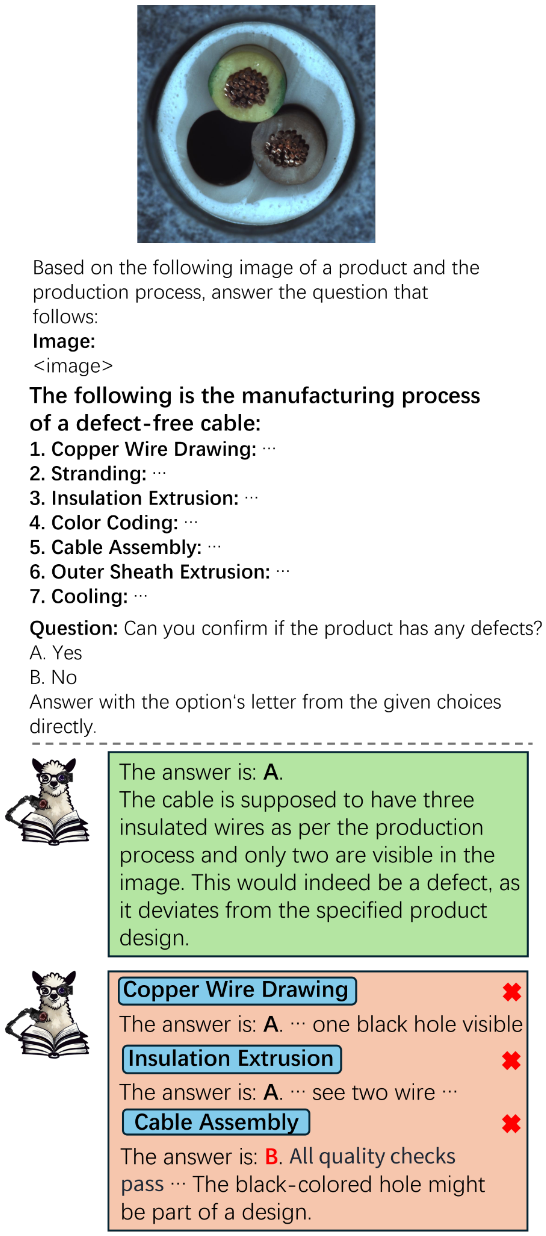


Figure A1. Ablation on steering the manufacturing process of the cable. The model is based on llava-1.6.

options are provided with additional instructions leading to a binary output of "A" or "B" for the convenience of calculating accuracy. For the LMM-based IAD methods, we keep their original prompts as much as possible. Specifically for AnomalyGPT [10], it has a hint before the question on MVTec AD [2].

The accuracy is calculated by retrieving a binary answer of "Yes" or "No" with keyword mapping. To evaluate the effect of context, the manufacturing process is directly added before the question. Some examples of the manufacturing process generated by LLM, which is used in evaluation on MVTec-AD are shown in Tab. A7.

C. Task details

We have manually annotated detailed IAD captions on VisA [41] and a part of RealIAD [30]. An example is shown in Fig. A3 with a human-annotated caption. When labeling the IAD image, we first describe the product in the image and its detailed features following the general image-captioning task. Then, the defect is described by its type, location, texture, color, and shape. For multiple defects, each defect will be described separately with its relative position to the image. Later, captions are processed by Large Language Models (LLMs) to construct different task-based instruction datasets.

The examples of four tasks used in InstructIAD and CoT-M are shown in Fig. A5.

D. Further results about EG-RoI

D.1. Binarized anomaly segmentation map

Triad uses the binarized outputs of anomaly segmentation vision experts in EG-RoI. We observe that the range of anomaly scores varies between different vision experts, which makes it difficult to integrate them in a unified manner. The conventional normalization methods in [4, 7, 12] are based on the entire test set, which is not feasible in real applications and requires a manually selected threshold for further binarization. Considering that vision experts only provide suspicious regions, we simply normalize the scores to the range 0-1 in one anomaly map and retain the scores that are larger than 0.9 on the anomaly map. As for threshold, Tab. A4 shows that although accuracy varies with different thresholds, Triad has robustness against the increase of false anomaly regions (Pixel-FPR) and achieves a good performance when the True Positive Rate is low.

D.2. Impact of the accuracy of anomaly maps

We calculate the accuracy of anomaly maps provided by vision experts for different thresholds on the MVTec AD test set in Tab. A5. Though selecting the best threshold, the performance of vision experts is typically poorer than Triad, and they also lack the ability to utilize extra context, such as the manufacturing process to improve their performance on new products.

E. Other Explains

E.1. Confidence Voting Mechanism

Confidence Voting Mechanism (CVM) is used in one-shot evaluation with the zero-shot model boosting the one-shot model in Traid (LLaVA-1.6) due to the lack of multi-image

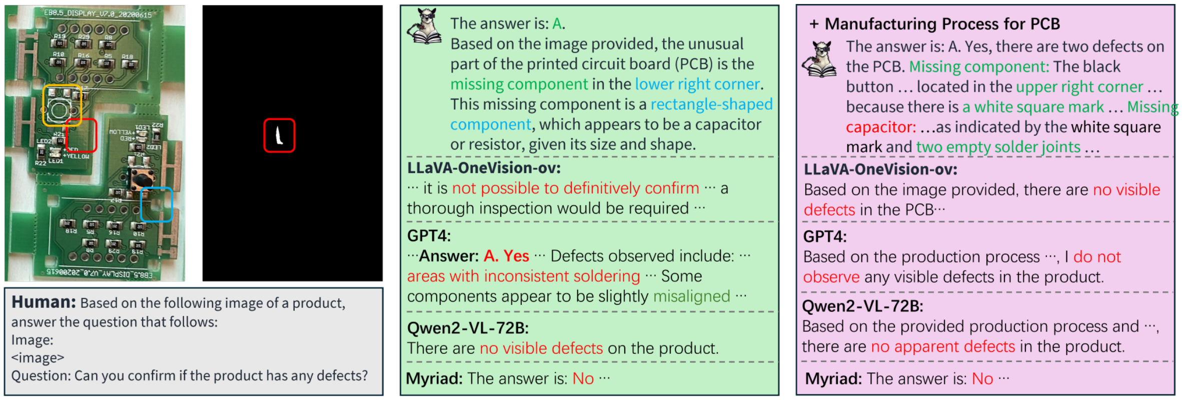


Figure A2. Qualitative results on another examples of PCB7 in PCB-Bank [33]. The comparison between SOTA LMMs and one of the LMM-based IAD methods is presented. Our chosen vision expert ([18]) makes mistakes. The bounding box used by Triad is plotted in red, while the ground truth is plotted in yellow. Another component noticed by Triad is highlighted by blue bounding boxes. Best view in **color**.

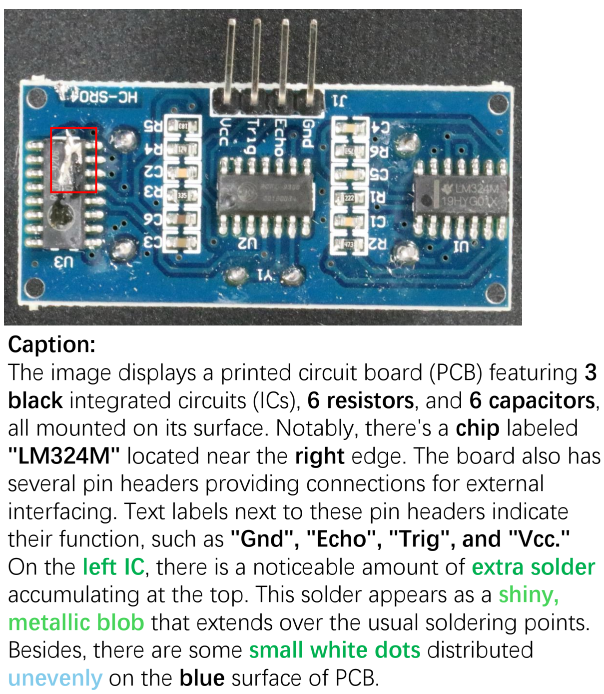
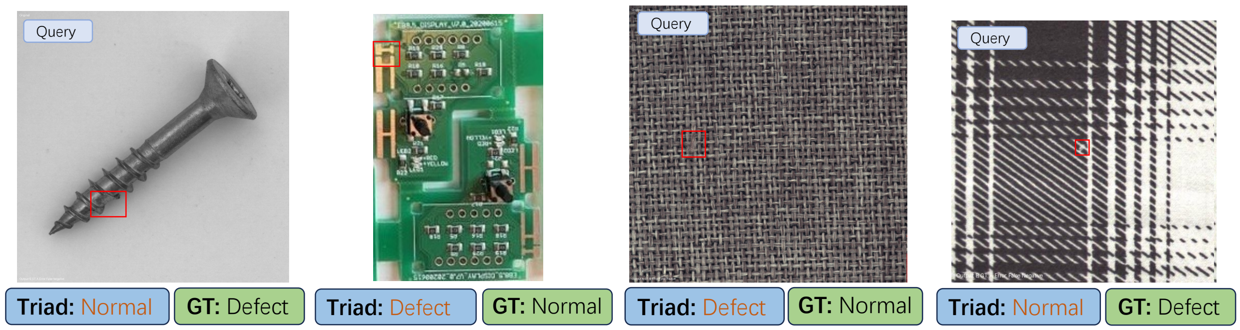


Figure A3. An example of attribute-rich caption in InstructIAD

Table A3. Comparison of 0-shot model and 1-shot model with or without CVM on Triad-lava-1.6.

model	1-shot base	+ MFG Proc.
0-shot model	85.0%	87.5% (2.5%↑)
1-shot model	87.8%	88.3% (0.5%↑)
1-shot model (CVM)	87.7%	88.4% (0.7%↑)

ability of LLaVA-1.6 [22]. In LMMs, references do not always bring benefits; instead, they could be misleading if the reference has some differences from the query image. Thus, the opinion of the zero-shot model is worth thinking about whether LMM is not strong enough for understanding multiple images. Confidence Voting Mechanism uses scores to collect the results of zero-shot model and one-shot model and give a combined result. As LMMs are bad at outputting confidence scores, the Confidence Voting Mechanism uses the option's word predict score from the language model and compares them by following rules: If two models reach a consensus, directly give the result; else if one model gives a negative opinion, which is saying the query is non-defective, its word predict score would be compared with the reference predict score by the same model. If the model is trusting the query more like a normal sample than the reference, it will be trusted and result in a non-defective judgment; otherwise, the opposite opinion will be adopted. The reason for comparing with the reference prediction score is that the space of predicted scores by the two models is not the same and unable to be normalized, so direct comparison makes no sense for the result. The effect of the Confidence Voting Mechanism is shown in Tab. A3.



(a) Screw (MVTec-AD) (b) PCB7 (PCB-Bank) (c) Carpet (MVTec-AD) (d) Grid cloth (WFDD)

Figure A4. Failure cases from MVTec-AD, WFDD, and PCB-Bank.

Referencing the image that are shown below, please answer the question:
Image:
<image>
Question: Could you identify any irregular features in the product displayed in the image?
A. Yes
B. No
Answer with the option's letter from the given choices directly.

Answer: A.

(a) anomaly detection

Referencing the image that are shown below, please answer the question:
Image:
<image>
Question: Describe the product in this image.

Answer: The image displays a printed circuit board (PCB) featuring 3 black integrated circuits (ICs), 6 resistors, and 6 capacitors, all mounted on its surface. Notably, there's a chip labeled "LM324M" located near the right edge. The board also has several pin headers providing connections for external interfacing. Text labels next to these pin headers indicate their function, such as "Gnd", "Echo", "Trig", and "Vcc." On the left IC, there is a noticeable amount of extra solder accumulating at the top ···

(b) attribute-level caption

Referencing the image that are shown below, please answer the question:
Image:
<image>
Question: Could you identify any irregular features in the product displayed in the image?
A. Yes
B. No

Answer: The answer is A.
The blue printed circuit board, identified as an HC-SR04 ultrasonic distance sensor module, exhibits **two** defects in the form of unwanted solder marks. The first defect is located on the surface of the **Integrated Circuit (IC)**···

(c) anomaly analysis

Based on the following image/image caption of a product and the manufacturing process, answer the question that follows:
Image:
<image> / <image caption>
The following is the manufacturing process of a defect-free PCB:
<MFG>
Question: Can you confirm if the product has any defects?
A. Yes
B. No

Answer: The answer is A.
Visual Selection Checklist for Defect-Free PCBs:
1. Substrate Condition:
 - Ensure there are no cracks, scratches, or uneven surfaces on the PCB substrate. Check: No defect ···

(d) CoT-M data organization

Figure A5. Tasks templates used in InstructIAD and CoT-M

Table A4. Impact of threshold in EG-RoI (LLaVA-1.6/MuSc).-0.1 means the top 10% regions with the lowest likelihood are marked, so does -0.2.

Threshold	0.9	0.8	0.7	0.6	0.5	0.4	0.3	0.2	0.1	-0.1	-0.2
base	85.0%	86.8%	85.5%	85.2%	84.3%	85.1%	83.9%	84.5%	83.8%	84.4%	84.9%
+MFG (LLM)	87.5%	87.8%	86.7%	86.5%	86.1%	85.8%	85.2%	85.3%	84.6%	84.9%	84.7%
Δ	2.5%↑	1.0%↑	1.2%↑	1.3%↑	1.8%↑	0.7%↑	1.3%↑	0.8%↑	0.8%↑	0.5%↑	0.2%↓
Pix-TPR (Expert)	32.1%	62.6%	80.1%	89.9%	95.8%	98.7%	99.6%	99.9%	100.0%	0.0%	0.1%
Pix-FPR (Expert)	0.6%	3.3%	8.3%	16.4%	28.3%	43.4%	62.0%	82.3%	96.7%	3.3%	17.8%

Table A5. Accuracy calculated by the anomaly map we used on MVTec AD. The (0shot, 1shot) test is conducted on Triad-ov.

Threshold	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	Best Acc	Triad (Base)	Triad (+MFG)
AnomalyClip [40]	73.3%	75.1%	76.8%	77.4%	78.4%	79.5%	80.3%	80.2%	**80.9%**	80.9%	(90.5%, 92.9%)	(91.6%, 93.0%)
MuSc [18]	72.3%	72.3%	72.3%	72.3%	72.3%	**86.3%**	66.1%	39.5%	28.5%	86.3%	(91.2%, 92.9%)	(92.6%, 94.1%)

Table A6. Prompt template used in evaluation. Manufacturing processes for each class of product are inputted as context.

Different methods	0shot prompt	+MFG. proc prompt
General LMMs (Traid on LLaVA-1.6)	*<image>* Are there any defects on the *<object_name>* in this image? A. Yes B. No Answer with the option's letter from the given choices directly.	*<image>* *<context>* Are there any defects on the *<object_name>* in this image? A. Yes B. No Answer with the option's letter from the given choices directly.
LLaVA-OneVision (Triad on LLaVA-OneVision)	Referencing the image that is shown below, please answer the question: Image: *<image>* Question: Are there any defects on the *<object_name>* in this image? A. Yes B. No Answer with the option's letter from the given choices directly.	Referencing the image and production process that are shown below, please answer the question: Image: *<image>* *<context>* Question: Are there any defects on the *<object_name>* in this image? A. Yes B. No Answer with the option's letter from the given choices directly.
Myriad	*<Image><ImageHere><\Image>* This image may be simulated by photo editing. According to IAD expert opinions and corresponding visual descriptions, find out if there are defects in this image.	*<Image><ImageHere><\Image>* *<context>* This image may be simulated by photo editing. According to IAD expert opinions and corresponding visual descriptions, find out if there are defects in this image.
AnomalyGPT	*</Img>* *<hint>* Is there any anomaly in the image?	*</Img>* *<context>* *<hint>* Is there any anomaly in the image?

Table A7. Examples of the manufacturing process generated by LLM about products in MVTec AD.

Class Name	MFG Proc.
Hazelnut	The following is the production process of the hazelnuts: 1. Harvesting: Hazelnuts are harvested from trees when they are ripe, usually in the late summer or early fall. Some variations in the size and color of the hazelnuts can occur based on the variety of tree and the growing conditions. This is not a defect, as it is normal for hazelnuts to have some variation in size and color. 2. Drying: After harvest, the hazelnuts are dried to reduce their moisture content. This helps to preserve the nuts and prevent spoilage. Some variation in the drying time and temperature can occur, which may cause slight differences in the color and texture of the hazelnuts. However, this is not a defect, as long as the hazelnuts have been dried to the correct moisture level. 3. Sorting: The dried hazelnuts are then sorted to remove any impurities, such as leaves, stems, and rocks. Some variation in the size and shape of the hazelnuts can occur during the sorting process. This is not a defect, as long as the hazelnuts are of a consistent size and shape suitable for the intended use. 4. Cracking: The sorted hazelnuts are then cracked to remove the shell. Some variation in the size and shape of the hazelnut kernels can occur during the cracking process. This is not a defect, as long as the kernels are intact and of a consistent size and shape. 5. Grading: The cracked hazelnuts are then graded based on their size and quality. Some variation in the size and quality of the kernels can occur. This is not a defect, as long as the hazelnuts meet the required standards for their intended use. 6. Packaging: The graded hazelnuts are then packaged for sale. Some variation in the packaging process can occur, which may cause slight differences in the appearance of the final product. However, this is not a defect, as long as the hazelnuts are properly protected and labeled.
Pill	The following is the production process of the pills: 1. Formulation: This is the first step where the active ingredients, excipients, and binders are mixed together in a specific ratio to form a powdered mixture. The differences that may occur in this step are related to the uniformity of the mixture, which can affect the dose of the active ingredient in each pill. However, these variations are usually not a defect as long as they fall within the acceptable range set by regulatory bodies. 2. Granulation: The powdered mixture is then moistened with a binder solution to form granules. The granules are then dried and screened to ensure uniformity in size. The differences that may arise in this step include variations in the size and shape of the granules, which can affect the flow properties of the material during compression. However, these differences are not considered defects as long as they do not significantly impact the uniformity or quality of the final product. 3. Compression: The dried granules are compressed into tablets using a tablet press. The differences that may occur in this step include variations in the weight, hardness, and thickness of the tablets, which can affect the dissolution rate and bioavailability of the active ingredient. However, these differences are not considered defects as long as they fall within the acceptable range set by regulatory bodies. 4. Coating: The compressed tablets are then coated with a film of polymer or sugar to improve their stability, appearance, and taste. The differences that may occur in this step include variations in the thickness and uniformity of the coating, which can affect the release profile of the active ingredient. However, these differences are not considered defects as long as they do not significantly impact the stability or efficacy of the final product.

Table A8. The manufacturing process of cables generated by LLM and GPT.

Class Name	MFG Proc.
Cable (GPT)	The following is the production process of the cables: 1. Copper Wire Drawing: - Start with large copper rods and draw them through a series of progressively smaller dies to create fine strands of copper wire. 2. Stranding: - Multiple fine copper wires are stranded together to form a single conductor for each wire. 3. Insulation Extrusion: - Extrude insulation material around each stranded copper conductor. The insulation material is heated and then extruded around the wire as it passes through a die. 4. Color Coding: - During the insulation process, the insulation is color-coded with different colors to differentiate the wires within the cable. 5. Cable Assembly: - The three insulated wires are then twisted or laid together to form a single cable. This step ensures that the wires are held together in a compact and organized manner. 6. Outer Sheath Extrusion: - An outer sheath is extruded around the assembled wires to provide additional protection and integrity to the cable. 7. Cooling: - The extruded cable is passed through a cooling bath or air-cooling system to solidify the insulation and outer sheath materials.
Cable (LLM)	The following is the production process of the cables: 1. Stripping: This is the process of removing the outer jacket of the cable to expose the inner conductors. Differences in the length or precision of the stripped area may occur, but they are not considered defects as long as they are within the tolerable range specified by industry standards. 2. Twisting: In this step, the conductors are twisted together to form a cable. The number of twists per inch and the tightness of the twist may vary slightly from one production to another, but these differences do not affect the functionality or quality of the cable. 3. Insulation: Each twisted pair of conductors is covered with an insulating material to prevent electrical interference. The thickness and uniformity of the insulation may differ slightly between productions, but these differences are not considered defects as long as they comply with industry standards. 4. Braiding: A braided shield is added around the insulated conductors to provide additional protection against electrical interference. The tightness and uniformity of the braid may differ slightly between productions, but these differences are not considered defects as long as they do not affect the cable's performance. 5. Jacketing: A final outer jacket is added to encase all the inner components of the cable. The thickness and uniformity of the jacket may differ slightly between productions, but these differences are not considered defects as long as they are within the tolerable range specified by industry standards.