# 试论计算机视觉在质检管理中的误判类型与校正策略

毛子豪
（昆明理工大学管理与经济学院，云南 昆明 650093）

**摘　要：**计算机视觉技术已成为智能制造质检环节的核心手段，但其在实际部署中面临的误判问题严重制约了检测系统的可靠性与落地价值。本文从统计决策与生产管理的双重视角出发，对计算机视觉质检系统的误判进行系统类型学划分，将其归纳为假阳性（类型Ⅰ错误）、假阴性（类型Ⅱ错误）两类基本形态，并分析两类错误在返工成本、质量风险与品牌声誉等维度上的非对称性后果；进一步引入动态误判概念，揭示光照、角度、遮挡等环境因素诱发的情境性偏差及其传导机制。在此基础上，本文从数据、模型与流程三个层面提出误判校正的技术策略与原理：数据层面通过难例挖掘、数据增强与领域自适应提升训练样本的覆盖度与鲁棒性；模型层面引入不确定性量化与置信度阈值的动态调节机制，使模型能够识别自身认知边界并输出可靠的决策置信度；流程层面构建人机协同审核机制，将人工复核作为误判的最后防线，形成“算法初筛—人工复核—反馈迭代”的闭环治理体系。研究表明，计算机视觉质检的误判治理是一个涉及数据质量、模型能力与流程设计的系统性工程，单一维度的优化难以从根本上解决误判问题，需要多策略协同发力。

**关键词：**计算机视觉；质量检测；误判；假阳性；假阴性；不确定性量化；人机协同
**中图分类号：**TP391.41　　**文献标识码：**A　　**文章编号：**3134-9986(2026)04-0010-03

## 一、误判的系统类型学划分

### （一）假阳性误判：将合格品标记为缺陷的类型I错误

假阳性误判（False Positive, FP），在统计假设检验的语境中对应类型Ⅰ错误（Type I Error），即“弃真”错误——将原本合格的产品错误地判定为缺陷品。在计算机视觉质检系统中，假阳性表现为算法在无缺陷的产品表面识别出不存在的缺陷，或将产品表面的正常纹理、反光、污渍等非缺陷特征误判为划痕、凹坑、裂纹等质量缺陷。

假阳性误判的生成机制具有多层次性。从数据层面看，训练集中正样本（合格品）的多样性不足是导致假阳性的重要原因。工业产品的表面纹理、材质特性与生产工艺存在天然的批次差异，如果训练数据未能充分覆盖这些正常变异，模型在遇到未见过的正常外观时容易将其识别为异常。例如，在金属表面质检中，不同批次钢材的表面粗糙度、氧化膜颜色与纹理方向存在细微差异，当模型仅在单一批次数据上训练时，新批次的正常表面可能因分布偏移而被误判为缺陷。从模型层面看，特征提取的粒度与决策边界的设定直接影响假阳性率。过于敏感的模型倾向于将任何与标准模板存在偏差的区域标记为缺陷，而缺乏足够正则化的模型则容易在决策边界附近产生不稳定的预测。从成像层面看，光照不均、镜头畸变、传感器噪声等硬件因素可能在合格品图像中引入伪缺陷特征，如金属表面的高光反射可能被误判为划痕，织物表面的正常褶皱可能被误判为破损。

假阳性误判对生产管理的直接影响是增加不必要的返工与报废成本。被误判为缺陷的合格品将进入返工流程或直接报废，导致原材料、人工与设备工时的浪费。在高精度制造行业，如半导体芯片、光学镜片与精密机械零件，假阳性的代价尤为高昂——一片被误判为缺陷的晶圆可能导致数千美元的直接损失，而频繁的假阳性还会降低产线的有效产出率，增加单位产品的制造成本。更为隐蔽的是，假阳性会侵蚀质检系统的可信度。当一线工人发现算法频繁“误杀”合格品时，可能对系统的判断产生怀疑，进而忽视系统的真实报警，形成“狼来了”效应——当真正的缺陷出现时，反而可能被当作误判而忽略。这种信任侵蚀在管理层面可能导致企业对计算机视觉质检系统的整体投资回报产生质疑，延缓智能制造的推进节奏。

假阳性误判的治理需要在检测灵敏度与误检容忍度之间寻求平衡。在传统机器视觉检测中，通常通过调整判定阈值来控制假阳性率——提高阈值可以减少误检，但同时可能降低对真实缺陷的检出能力。在深度学习方法中，假阳性的控制更为复杂，需要通过数据增强、难例挖掘与模型正则化等综合手段实现。值得注意的是，假阳性率并非越低越好，过度追求低假阳性可能导致模型变得过于“宽松”，从而使假阴性率上升。两类错误之间存在此消彼长的权衡关系，这

**作者简介：**毛子豪，博士，副教授，研究方向为智能质检与视觉管理。
**本刊网址：**https://www.esepress.org/

正是误判治理的核心难点所在。

**（二）假阴性误判：将缺陷品放行的类型II错误**

假阴性误判（False Negative, FN），对应统计假设检验中的类型Ⅱ错误（Type II Error），即“取伪”错误——将存在缺陷的产品错误地判定为合格品并予以放行。在计算机视觉质检系统中，假阴性表现为算法未能识别出产品表面真实存在的划痕、裂纹、气泡、缺料等缺陷，导致不合格品流入下游工序或最终市场。

假阴性误判的成因与假阳性存在本质差异。从缺陷特征来看，微小缺陷、低对比度缺陷与背景纹理相似的缺陷是假阴性的高发区域。例如，钢材表面的细微划痕在特定光照下对比度极低，与人眼难以察觉的缺陷类似，深度学习模型也容易因特征信号微弱而漏检。又如，印刷电路板上的微小焊点虚焊，其外观特征与正常焊点高度相似，仅在细节上存在差异，模型若缺乏足够的细粒度特征提取能力，极易将其误判为合格品。从数据层面看，缺陷样本的稀缺性与类别不平衡是导致假阴性的根本原因。在实际生产中，缺陷品的比例通常远低于合格品，某些罕见缺陷的样本量可能仅有数十张，模型在这类缺陷上的学习严重不足，导致泛化能力差、漏检率高。从模型架构来看，感受野大小、特征融合方式与注意力机制的设计直接影响模型对不同尺度缺陷的检测能力。过小的感受野可能遗漏大尺度的全局性缺陷，而过大的感受野则可能丢失微小缺陷的细节信息。

假阴性误判的后果通常比假阳性更为严重，因为它直接关系到产品质量与用户安全。缺陷品一旦流出工厂，可能在下游装配中引发连锁故障，或在终端使用中导致安全事故。在汽车零部件、医疗器械与航空航天等安全关键领域，假阴性的代价可能是灾难性的——一个被漏检的裂纹可能导致零部件在使用中断裂，造成人身伤害与巨额赔偿。即使在普通消费品领域，流出的缺陷品也会引发客户投诉、退货与品牌声誉损害，增加企业的售后成本与市场信任危机。从质量管理的角度看，假阴性意味着质量控制体系的失效，是六西格玛等质量管理方法论重点防控的对象。

假阴性的治理在工业实践中通常被赋予更高的优先级。许多企业在部署计算机视觉质检系统时，明确要求“零漏检”或“漏检率趋近于零”，即使这意味着较高的假阳性率。这种策略选择反映了两类错误的非对称性成本——漏检一个缺陷品的代价远高于误杀一个合格品的代价。然而，“零漏检”目标在技术上往往难以实现，过度追求低假阴性可能导致假阳性率失控，反而降低系统的整体可用性。因此，如何在严格控制假阴性的同时将假阳性维持在可接受范围内，是计算机视觉质检系统设计的核心挑战。

**（三）数据层面的校正：难例挖掘、数据增强与领域自适应**

数据是深度学习模型性能的根基，误判常源于训练数据的缺陷。数据层面策略旨在提升训练数据的质量、多样性与域适应性，核心方法包括难例挖掘、数据增强和领域自适应。

难例挖掘通过定向收集模型易误判的“困难样本”（如缺陷微弱或与正常样本高度相似的样本），重新标注后加入训练集，针对性地强化模型的薄弱环节。该过程通常迭代进行：用初始模型在验证集或产线数据上找出误判样本，经人工复核并重训，直至性能达标。实践中，持续的难例挖掘可将误判率从5%－10%降至1%以下，关键在于建立高效的标注与迭代流程。

数据增强通过对训练样本做几何变换、颜色变换、噪声添加等基本操作扩充数据，提升模型鲁棒性。更先进的方法如生成对抗网络、扩散模型和基于物理渲染的合成数据，可生成逼真的罕见缺陷样本，有效改善对小众缺陷的检出能力。增强需遵循“真实性”原则，避免引入不符合实际分布的样本。

领域自适应应对训练数据（源域）与部署数据（目标域）间的分布差异。无监督领域自适应无需目标域标注，通过对抗训练、最优传输或伪标签等方法对齐特征分布，实现知识迁移。该技术显著降低了跨产线模型迁移的数据采集与标注成本。

这三种策略互补协同：难例挖掘提升决策边界精度，数据增强扩充数据覆盖，领域自适应解决域偏移。实际应用需根据误判类型选择组合，例如高假阴性应侧重难例挖掘与生成式增强，动态误判则需加强环境模拟与域自适应。数据层面的校正是一切后续优化的基础。

**（四）模型层面的校正：不确定性量化与置信度阈值的动态调节**

该策略的核心是让模型识别自身预测的可靠程度，并据此动态决策，避免在不确定状态下武断判断。

深度学习的不确定性分为认知不确定性（因训练数据覆盖不足所致，可随数据增加而降低）和偶然不确定性（源于图像噪声或模糊，数据固有）。前者常对应未见过的缺陷类型，易致假阴性；后者常对应成像质量差的样本，易致假阳性或假阴性。

量化方法包括贝叶斯神经网络（精度高但计算繁重）、MC Dropout（推理时多次随机失活并统计预测方差，实用性好）、深度集成（多模型集成，效果好但

成本高）和证据深度学习（单次前向传播即可估计不确定性，兼具效率与准确性）。

在量化基础上，动态调节置信度阈值成为核心机制：低不确定性样本可用宽松阈值提高效率，高不确定性样本则用更严阈值或转交人工复核。也可结合误判成本调节，如在安全场景中降低缺陷阈值以减少漏检，增加的人工审核样本作为补偿。

此外，模型常存在“过度自信”问题，需进行置信度校准，使输出概率与真实准确率一致。常用方法包括温度缩放、等渗回归等，使校准后的置信度更可靠，为动态阈值与人机协同提供依据。

模型层面校正体现的是从“追求绝对准确”向“承认不确定性并智慧应对”的转变。在复杂环境中，更务实的做法是让系统在能力范围内自主判断，超出边界时主动求助人工，这正是质检系统走向成熟的关键标志。

## 二、结束语

随着基础模型与多模态大模型技术的发展，计算机视觉质检的误判治理将迎来新的机遇。大规模预训练模型所具备的强大泛化能力与零样本学习能力，有望显著降低对特定领域标注数据的依赖，减少因数据覆盖不足导致的误判；多模态融合技术则可以结合视觉、红外、声学等多种传感信息，提升对复杂缺陷的判别能力，降低单一模态下的不确定性。同时，可解释人工智能（XAI）技术的发展将使模型的误判原因更加透明可追溯，为人机协同与模型改进提供更精准的指导。然而，技术的进步并不能消除误判本身——在日益复杂的制造环境中，新的缺陷类型与新的环境挑战将不断涌现，误判治理将是一个持续迭代、永无止境的过程。对于制造企业而言，建立系统化的误判治理体系与持续改进机制，比追求某一时刻的“零误判”目标更具现实意义与长期价值。

## 参考文献：


[1] 表面缺陷视觉检测的深度学习技术 [M].武汉：武汉大学出版社，2024.

[2]机器视觉技术及应用（第 2 版）[M].北京：高等教育出版社，2023.

[3] 庞记成.基于双分支结构的小样本印刷品表面缺陷检测方法研究 [D].成都：成都理工大学，2025.

[4] 马冬梅，朱佳浩.面向热轧带钢表面缺陷检测的 YOLOv5 算法优化分析 [J].制造技术与机床，2024（6）：153-160.

[5] 基于深度学习的表面缺陷检测方法综述 [J].自动化学报，2021，47（12）：2785-2797.

[6] 基于计算机视觉的工业金属表面缺陷检测综述 [J].自动化学报，2024，50（3）：545-562.

[7] 基于深度学习的工业缺陷检测研究进展 [J].计算机科学，2024，51（10）：261-270.


# On Misjudgment Types and Correction Strategies of Computer Vision in Quality Inspection Management


Mao Zihao

*(Faculty of Management and Economics, Kunming University of Science and Technology, Kunming 650093)*



**Abstract:** Computer vision technology has become a core instrument in the quality inspection link of intelligent manufacturing, yet the misjudgment problem in its practical deployment severely restricts the reliability and deployment value of inspection systems. From the dual perspectives of statistical decision and production management, this paper conducts a systematic typological classification of misjudgments in computer vision quality inspection systems, summarizing them into two basic forms: false positives (Type I error) and false negatives (Type II error), and analyzes the asymmetric consequences of the two types of errors in dimensions such as rework cost, quality risk and brand reputation. The paper further introduces the concept of dynamic misjudgment, revealing the situational deviations induced by environmental factors such as illumination, angle and occlusion and their transmission mechanisms. On this basis, the paper proposes technical strategies and principles for misjudgment correction from three levels: data, model and process. At the data level, hard example mining, data augmentation and domain adaptation are used to improve the coverage and robustness of training samples. At the model level, uncertainty quantification and dynamic adjustment of confidence thresholds are introduced to enable the model to recognize its own cognitive boundary and output reliable decision confidence. At the process level, a human-machine collaborative review mechanism is constructed as the last line of defense against misjudgments, forming a closed-loop governance system of "algorithm initial screening—manual review—feedback iteration". The study shows that misjudgment governance in computer vision quality inspection is a systematic project involving data quality, model capability and process design, and optimization in a single dimension cannot fundamentally solve the misjudgment problem, requiring coordinated efforts of multiple strategies.



**Keywords:** computer vision; quality inspection; misjudgment; false positive; false negative; uncertainty quantification; human-machine collaboration