

技术交底书-工艺类

发明人姓名：

联系方式：

1. 发明创造名称

一种基于多阶段深度学习模型的智能 CAD 图像审图匹配系统及其方法

2. 背景技术和现有技术的缺陷与不足

在工业制造、建筑设计、电子电路设计等多个领域，CAD（计算机辅助设计）技术已经成为不可或缺的工具。CAD 文件用于设计、绘图、分析、仿真和制造等环节，确保从设计概念到实际产品的精确呈现。然而，随着工程项目的复杂性增加，CAD 文件之间的版本管理、版本比对和设计一致性验证成为一项关键任务。尤其是在涉及多个设计团队或跨领域协作时，CAD 文件的比对和审查变得更加复杂且至关重要。

2.1. 现有技术概述

传统的 CAD 图纸比对和审查方法主要依赖于手工操作或简单的几何对比。手工审查通常需要经验丰富的工程师或设计师进行，这种方法不仅效率低下，而且极易出现漏检或误判的情况。几何对比方法虽然引入了计算机辅助，但多局限于基本的形状相似度计算，难以应对复杂的 CAD 图纸变化。这些方法在处理大规模、多版本、多源的 CAD 文件时显得力不从心，难以满足现代工业和工程的需求。

近年来，随着计算机视觉技术的发展，基于图像处理的自动化 CAD 图纸比对方法逐渐受到关注。这些方法利用图像处理算法、特征提取技术和机器学习模型，对不同版本的 CAD 图纸进行比对，以实现自动化和智能化。然而，这些方法在实际应用中仍然存在显著的不足：

- 单阶段图像匹配的局限性：**现有的图像匹配方法多采用单阶段模型，通常只考虑图像的全局特征或局部特征。这种方法在面对复杂 CAD 图纸时容易受到形变、尺度变化、旋转等因素的影响，导致匹配精度不足。此外，单阶段方法难以处理 CAD 图纸中不同目标的细粒度特征，尤其是在多目标、多类别的场景下。
- 处理超大尺寸图像的挑战：**CAD 图纸往往具有超高分辨率和超大尺寸，现有的图像处理算法和模型难以直接应用。在大尺寸图像的目标检测中，传统的全局检测方法无法有效兼顾精度和效率，而单纯的局部检测方法（如滑窗检测）在拼合和全局处理时，容易引入冗余检测和漏检问题，进一步影响检测和匹配的准确性。
- 多类别不平衡问题：**在 CAD 图纸中，不同类别的目标通常具有高度不均衡的分布。例如，在一个复杂的电子电路设计图纸中，某些元件（如电容、电阻）的出现频率远高于其他元件。现有的图像处理和机器学习方法难以在样本类别分布严重不均衡的情况下，维持较高的匹配和分类精度。
- 缺乏细粒度特征匹配能力：**CAD 图纸中包含大量细节信息，如线条粗细、标注位置、字体风格等，现有方法往往仅关注几何特征或形状特征，缺乏对这些细节的有效处理能力。因此，无法在细节差异较大的场景下准确判断图纸的一致性。
- 处理异构数据的能力不足：**在实际工程应用中，CAD 文件的格式多样，既有矢量格式（如 DWG、DXF），也有图像格式（如 PNG、JPG），甚至还包括 PDF 等文档格式。现有技术通常只能处理其中一种或几种格式，缺乏对多种格式的统一处理能力，影响了自动化审图系统的适用性。

2.2. 背景技术的缺陷与不足

基于上述分析，现有技术在 CAD 图纸的自动化比对和审查中存在以下主要缺陷和不足：

1. 全局与局部特征的有效结合不足：现有方法在处理 CAD 图纸时，难以同时考虑全局结构和局部细节特征，导致匹配结果不够精确。尤其是在图纸包含复杂结构和细节变化时，单一特征的使用很难获得令人满意的匹配结果。
2. 大尺寸图像的处理效率低：由于 CAD 图纸通常为超大尺寸，现有检测算法在处理此类图像时，计算资源需求高、处理时间长，同时还容易产生误检和漏检。如何在保持高精度的同时提高处理效率，仍是现有技术的一个瓶颈。
3. 样本不均衡问题严重：在多目标检测和匹配中，样本类别的不均衡性显著影响了模型的性能。现有技术缺乏有效的策略来解决样本不均衡问题，导致在稀有类别的检测和匹配上存在明显的精度损失。
4. 缺乏适应多源数据的通用性：现有自动化审图方法多局限于单一数据源或格式，缺乏对多种异构数据（如 PDF、图像、矢量图）的统一处理能力，限制了技术的应用范围。
5. 难以实现端到端的智能化审图：目前的技术方案多为工具化产品，难以实现从数据处理、检测到匹配和分析的全流程自动化，智能化程度有限。特别是在跨领域应用时，现有技术无法满足多样化和个性化的需求。

3. 具体的技术方案描述

本发明提出了一种基于多阶段深度学习模型的智能 CAD 图像审图匹配系统，旨在通过多种先进技术，实现对任意两幅或多幅图像或 PDF 格式存储的任意尺寸 CAD 文件的智能匹配和判断。以下为本发明的完整技术方案描述，重点阐述具有独创性的部分，并对关键步骤进行详细说明。

3.1. 系统整体架构概述

系统整体架构包括以下几个主要模块：

1. 数据预处理与复杂数据集构建：对 CAD 图像进行预处理，构建平衡且具有多样性的训练数据集。
2. 目标检测模型：使用 YOLOv5 模型进行多目标检测，结合滑窗检测和全局非极大值抑制（NMS）技术，处理超大尺寸图像。
3. 目标匹配模型：采用双头网络结构，实现目标的精确位置匹配和类别匹配。
4. 模型推理与匹配判断：通过多种损失函数优化特征提取与匹配决策，使用余弦距离和 F1 范数进行最终匹配判断。

3.2. 数据预处理与复杂数据集构建

数据预处理是系统的重要基础，直接影响模型的鲁棒性和泛化能力。针对 CAD 图像的特点，本发明提出了以下处理方法：

1. CLAHE（自适应直方图均衡化）：通过增强图像的局部对比度，突出图像中的细节特征。
2. 数据增强：包括随机旋转、平移、缩放、裁剪、颜色抖动等多种手段，生成多样化的训练数据。
3. 样本均衡与复杂数据集构建：通过 SMOTE（合成少数类过采样技术）和 RandomUnderSampler（随机欠采样技术）对数据进行平衡，避免模型训练时受类别不均衡的影响。

通过上述步骤，系统生成了一个高质量、平衡的数据集，为后续模型的训练提供了坚实基础。

3.3. 目标检测模型训练

系统使用 YOLOv5 模型对 CAD 图像进行多目标检测。在处理超大尺寸图像时，采用了一种创新的多阶段检测方法：

1. 滑窗检测：将超大尺寸的 CAD 图像分割为多个小窗，每个小窗以合适尺寸输入 YOLOv5 模型进行检测。检测结果包括目标位置和置信度分数。
2. 局部 NMS 与全局 NMS：首先在每个滑窗内进行局部 NMS，抑制窗口内的冗余检测。然后，在全局范围内合并所有滑窗的检测结果，并进行全局 NMS，去除重叠目标。具体解释如下：

假设图像大小为 $W \times H$ ，滑窗大小为 $w \times h$ ，滑动步长为 s_x 和 s_y 。滑窗在图像上按以下公式进行移动：

$$x_i = i \cdot s_x, y_j = j \cdot s_y$$

其中 i, j 为整数，使得 $0 \leq x_i \leq W - w$ 和 $0 \leq y_j \leq H - h$ 。对于每个滑窗的检测结果 D_k ，通过局部 NMS 处理后，再将其在全局坐标系下进行转换，最终通过全局 NMS 进行融合，得到全局检测结果 D_{final} 。

经过目标检测中创新性地结合了滑窗检测与全局 NMS 技术，确保了对超大图像中目标的精准识别。滑窗检测使模型能够处理超大图像，全局 NMS 则避免了重叠检测带来的误判。

3.4. 目标匹配模型训练与推理

目标匹配模型采用双头网络结构，通过结合位置匹配和类别匹配，提高模型的识别和匹配能力。匹配过程中使用多种损失函数，具体步骤如下：

1. 双头网络架构：模型结构包括位置匹配分支和类别匹配分支。输入图像经过卷积层提取特征后，分别输出匹配特征向量和类别预测。
2. 三元组损失 (Triplet Loss with Hard Negative Mining)：用于位置匹配，通过余弦距离度量样本之间的相似性，并结合难例挖掘策略，优化特征向量，使得正样本之间距离最小，负样本之间距离最大：

$$L_{triplet} = \max(0, \cos(f_a, f_n) - \cos(f_a, f_p) + \text{margin})$$

3. ArcFace 损失：用于类别匹配，通过引入角度间隔，增强同类样本在高维空间中的聚集性，优化类别预测的准确性：

$$L_{ArcFace} = -\frac{1}{N} \sum_{i=1}^N \log \frac{e^{s(\cos(\theta_{y_i} + m))}}{e^{s(\cos(\theta_{y_i} + m))} + \sum_{j \neq y_i} e^{s \cos \theta_j}}$$

4. NTXent 损失：在对比学习阶段使用，优化特征分布，增强不同类别之间的区分度。此损失通过温度缩放机制调整特征分布，使相似样本之间的距离尽可能接近：

$$L_{NTXent} = -\log \frac{e^{\frac{\text{sim}(z_i, z_j)}{\tau}}}{\sum_{k=1, k \neq i}^{2N} e^{\frac{\text{sim}(z_i, z_j)}{\tau}}}$$

5. 类别敏感交叉熵损失：针对类别不平衡问题，在类别匹配阶段应用，确保模型在类别分布不均衡情况下依然具有较好的鲁棒性：

$$L_{CE} = - \sum_i w_{y_i} \log p_{y_i}$$

6. 知识蒸馏损失：在学生网络训练阶段，通过最终输出 logits 和特征向量 features 的知识蒸馏，将教师网络的知识转移给学生网络，实现模型的轻量化，同时保持高性能：

$$L_{distill} = \frac{\lambda_{logits}}{T^2} KL\left(\sigma \frac{z_{teacher}}{T}, \sigma \frac{z_{student}}{T}\right) + \lambda_{features} \|f^{teacher} - f^{student}\|^2$$

7. BYOL 自监督学习损失：用于教师网络的特征提取，通过自监督学习增强无标签数据的表达能力，提升模型的泛化性能：

$$L_{BYOL} = \frac{1}{2} \left[\|\hat{q}\theta(z_{online1}) - sg(z_{target2})\|^2 + \|\hat{q}\theta(z_{online2}) - sg(z_{target1})\|_2^2 \right]$$

8. 网络参数 L2 正则化损失：用于抑制复杂网络在复杂任务训练过程中的过拟合问题：

$$L_{L2} = \lambda \sum_k \|\theta_k\|^2$$

最终我们得到教师模型使用的综合性损失为：

$$L_{teacher} = \alpha_1 L_{triplet} + \alpha_2 L_{Arcface} + \alpha_3 L_{NTXent} + \alpha_4 L_{CE} + \alpha_5 L_{BYOL} + \alpha_6 L_{L2}$$

得到学生模型使用的综合性损失为：

$$L_{student} = \alpha_1 L_{triplet} + \alpha_2 L_{Arcface} + \alpha_3 L_{NTXent} + \alpha_4 L_{CE} + \alpha_5 L_{distill} + \alpha_6 L_{L2}$$

通过多种损失的综合作用，显著提升了匹配模型的精度与稳定性。三元组损失和 ArcFace 损失确保了位置匹配与类别匹配的精度，而 BYOL 和 NTXent 损失则增强了特征的鲁棒性与通用性。知识蒸馏损失有效实现了教师-学生网络之间的知识传递，使得轻量化模型具备高性能。

3.5. 模型推理与匹配判断

在推理阶段，系统将提取的特征向量输入到匹配判断模块，使用余弦距离和 F1 范数进行最终的匹配判断。

1. 余弦距离计算：系统使用余弦距离作为匹配度量标准，判断两个目标的相似度：

$$D_{cos} = 1 - \frac{f_1 \cdot f_2}{\|f_1\| \cdot \|f_2\|}$$

2. 最优匹配阈值的计算：为了确定最佳的匹配阈值，系统通过计算正负样本的 F1 范数，选取最优阈值，确保匹配结果的准确性与稳定性：

$$F_1 = \frac{2 \cdot P \cdot R}{P + R}$$

3.6. 本发明的独创性和优势

本发明通过创新性地结合滑动检测与全局 NMS、双头网络架构、多损失函数组合，以及高效的推理和匹配判断，解决了现有技术在 CAD 图像审图匹配中的多个难题，为实际工业和工程应用提供了一种高效、智能的解决方案。

4. 本发明创造的优点

针对上述现有技术的缺陷与不足，本发明具有如下优点：

1. 多阶段检测与匹配机制：通过滑窗检测、局部 NMS 与全局 NMS 的结合，解决了超大尺寸 CAD 图像的处理难题，实现了高精度的目标检测与匹配，避免了传统方法中的冗余检测与漏检问题。
2. 双头网络结构的高效匹配：本发明采用双头网络结构，将位置匹配与类别匹配独立处理，使得类别信息不会干扰位置匹配，同时利用类别信息辅助匹配决策，提升了系统的匹配精度和鲁棒性。
3. 多损失函数的组合优化：通过多种损失函数（如三元组损失、自监督学习损失、蒸馏损失、ArcFace 损失等）的组合使用，全面提升模型的特征表达能力、匹配精度和分类性能，特别是在处理复杂样本和类别不平衡时表现尤为出色。
4. 高效的知识蒸馏与模型轻量化：通过知识蒸馏技术，将教师网络的高性能特征传递给学生网络，实现了模型的轻量化设计，使得系统能够在资源受限的环境中依然保持高效的推理性能。
5. F1 范数指导的最优匹配判断：本发明采用 F1 范数计算最优匹配阈值，确保匹配结果的准确性和稳定性，特别是在召回率与准确率之间取得了良好的平衡。
6. 通用性与扩展性：系统支持多种数据格式（如 PDF、图像、矢量图）的处理，适应多源异构数据的需求，具有良好的通用性与扩展性，适用于多种工业和工程应用场景。

5. 具体实施方式和附图

实施方式一：双头匹配网络的训练过程

本发明中的双头匹配网络采用教师-学生架构，结合多种损失函数，通过知识蒸馏策略提升学生模型的性能。具体实施方式如附图 1 所示。

在训练过程中，教师网络与学生网络分别使用以下损失函数进行训练：

1. 教师网络损失函数：
 - 带有难例挖掘的三元组损失：通过优化特征向量的相似性，确保同类样本间距离更近，不同类样本间距离更远。
 - BYOL 自监督损失，最小化网络模型对同一个源图像派生出的两个不同的 transform 图的均方差损失，强化网络学习图像的内在特征的能力。
 - ArcFace 损失：通过引入角度间隔，增强同类样本在高维空间中的聚集性。
 - NTXent 损失：在对比学习中，通过温度缩放机制优化特征分布。
 - 类别敏感交叉熵损失：引入类别权重，增强模型在不平衡样本情况下的鲁棒性。
 - L2 正则化损失：通过对模型参数进行 L2 正则化，抑制过拟合。
2. 学生网络损失函数：
 - 损失同教师网络，但不使用 BYOL 自监督损失，增加蒸馏损失。
 - 蒸馏损失：包括 logits 蒸馏和特征向量蒸馏，确保学生网络在保持高效轻量化的同时获得教师网络的知识。

实施方式二：CAD 超大图的整体匹配流程

针对超大尺寸 CAD 图像的匹配任务，本发明设计了一套完整的流程，能够高效、准确地对超大图中的元件进行检测和匹配。具体实施方式如附图 2 所示。

该流程分为以下几个关键步骤：

1. 超大图获取：首先获取两个超大尺寸的 CAD 图像。
2. 双图像尺寸对齐：通过 letterbox 填充方式对两幅图像进行尺寸对齐，以确保后续的匹配步骤能够在统一的尺度下进行。
3. 滑窗检测：对齐后的图像通过滑窗方式进行分块检测，确保每个局部区域的目标能够被有效检测到。
4. 全局坐标恢复与 NMS：对滑窗检测得到的局部目标进行全局坐标恢复，并通过全局非极大值抑制（NMS）去除冗余的检测框。
5. 基于匈牙利算法的双图像元件位置匹配：通过加权匈牙利算法，对两幅图像中的元件位置进行匹配，得到可能的匹配对。
6. 基于余弦距离的匹配网络元件匹配：利用双头匹配网络对上述匹配对进行进一步验证，通过余弦距离计算匹配分数。
7. 相似度分析与差异点分析：对所有匹配对进行相似度分析和差异点分析，最终输出结果并生成 JSON 文件。

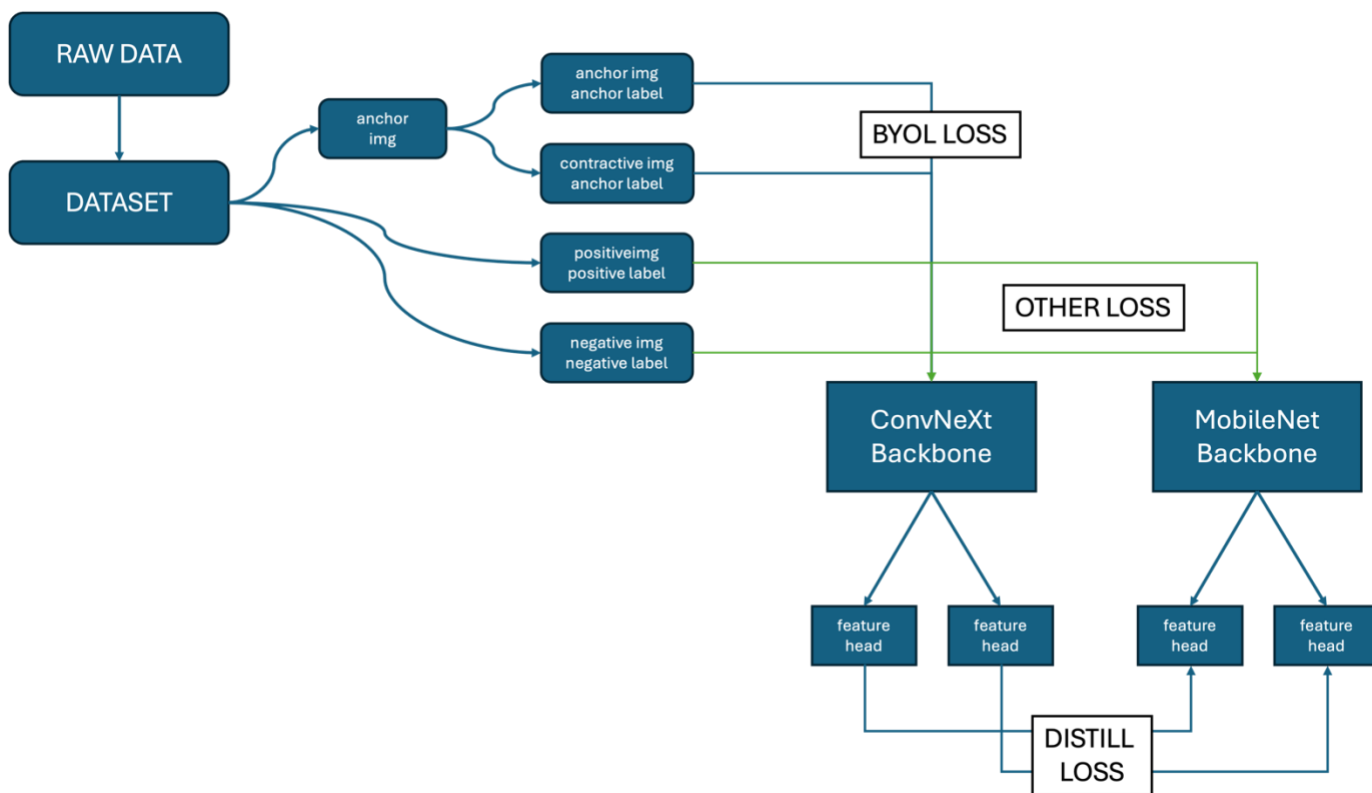


图 1 双头图像匹配网络的结构与训练

