

人机协同视域下中职数字媒体作品评价智能体的构建与应用

陶琴

王林“天山英才”名师工作室项目, 中国·新疆 乌鲁木齐 830000

摘要: 中职数字媒体专业图形图像处理课程涉及大量实作品, 传统人工批改长期受困于耗时多、标准波动、反馈粗略及反馈滞后等问题。本研究以人机协同评价理论为指导, 面向中职数字媒体作品批改场景构建了评价智能体, 设计了覆盖修瑕、抠像、调色、材质合成等维度的分类型评价指标体系, 借助提示词工程、知识库配置与 workflow 编排完成智能体搭建, 并构建了“AI 初评→教师复核→学生修改→AI 复评”的闭环评价流程。一学期的教学实践表明, 该智能体推动作品评价从教师单向批改转向学生自主参与、AI 即时反馈与教师针对性指导的协同模式, 反馈周期由一周左右缩减至当天, 评价标准一致性系数由 0.72 升至 0.89, 学生对反馈及时性的满意度由 58% 升至 89%, 作品及格率由 87% 升至 94%。本研究为人工智能辅助教育评价在中职领域的落地提供了可复制的实践经验。

关键词: 评价智能体; 中职教育; 数字媒体作品; 人机协同评价; 提示词工程

Construction and Application of Intelligent Agent for Digital Media Works Evaluation in Secondary Vocational Education from the Perspective of Human-Machine Collaboration

Tao Qin

Wang Lin "Tianshan Talents" Master Teacher Studio Project, China Xinjiang Urumqi 830000

Abstract: The course of Graphics and Image Processing for digital media majors in secondary vocational education involves a large number of practical works. Traditional manual evaluation has long been plagued by excessive time consumption, inconsistent evaluation criteria, superficial feedback and delayed responses. Guided by the theory of human-machine collaborative evaluation, this study constructs an intelligent evaluation agent for the grading of digital media works in secondary vocational education. A categorized evaluation indicator system covering retouching, matting, color grading and material compositing is established. The intelligent agent is developed through prompt engineering, knowledge base configuration and workflow arrangement, and a closed-loop evaluation process of "initial AI evaluation → teacher review → student revision → re-evaluation by AI" is formulated. The teaching practice over one semester shows that this intelligent agent transforms the traditional one-way evaluation by teachers into a collaborative mode featuring student independent participation, real-time AI feedback and targeted teacher guidance. The feedback cycle is shortened from about one week to the same day. The consistency coefficient of evaluation criteria rises from 0.72 to 0.89. Students' satisfaction with feedback timeliness increases from 58% to 89%, and the pass rate of works goes up from 87% to 94%. This study provides replicable practical experience for the implementation of artificial intelligence-assisted educational evaluation in secondary vocational education.

Keywords: Intelligent evaluation agent; Secondary vocational education; Digital media works; Human-machine collaborative evaluation; Prompt engineering

1 问题的提出

数字媒体技术应用是中职计算机类专业群中备受青睐的专业方向, 其课程体系以实践性见长, 学生须通过大量动手操作来习得图像处理、视频剪辑等技能^[1]。以图形图像处理课程为例, 每学期学生需提交十余次实操作业, 教师则需逐份审阅并给出评价。然而, 这一看似常规的教学

环节在实际运行中却暴露出多重矛盾。

近年来, 大语言模型与视觉理解模型的技术融合, 使得对图像作品进行自动化评价逐渐成为现实。本研究正是基于这一技术背景, 将评价智能体引入中职数字媒体作品批改场景, 探索人机协同评价的有效路径, 以期在减轻教师批改压力的同时, 提升评价标准的一致性、缩短反馈周

期、增强反馈的实效性。

2 文献综述与研究设计

2.1 人机协同评价的理论基础

人机协同评价主张人类评价者与人工智能系统在评价过程中各司其职、优势互补^[1]。从评价方式的历史演进看，由纯人工评价到计算机辅助评价再到智能评价，人机协同评价是当前最能兼顾效率与专业性的评价模式^[2]。人工智能在评价中的优势集中体现在三个方面：效率方面，一份作品的 AI 初评约 30 秒即可完成，而人工初评通常需 2 至 3 分钟；稳定性方面，评价标准一旦确定便不会发生漂移，不受时间、情绪等因素干扰；客观性方面，能够规避人类评价者可能存在的偏好或偏见^[2]。但人工智能同样存在明显短板——专业判断深度不足、情境理解能力有限、情感互动缺失，这些不足恰恰划定了教师在协同评价中不可逾越的边界^[3]。

2.2 人工智能辅助教育评价的研究现状

人工智能辅助教育评价已在多个学科积累了实践案例。语文学科中，智能系统可自动评分并识别作文中的错别字；数学学科中，系统可评判解题步骤并追踪思维过程；英语学科中，口语评测系统能够评估发音的准确度和流利度^[4]。在艺术设计领域，虽有研究尝试用深度学习模型评价绘画作品的美学质量，但多数尚处于技术可行性验证阶段，真正走进课堂的案例寥寥无几。中职教育领域的作品评价具有鲜明的实践性和技能导向特征，评价不仅关注最终结果，还需审视操作过程的规范性及技能掌握的熟练度，这对智能评价工具的情境适配能力提出了更高要求。

2.3 研究设计

研究对象是我教的两个中职数媒班，在一学期内围绕四件事逐项推进：先搭评价指标体系，再把评价智能体做成一个能用的工具，然后设计一套教师和 AI 配合着批作业的流程，最后在真实课堂上跑一整个学期看效果。方法上我主要用了三件套：一文献研究法；二是行动研究法，边教边用边调细节，前前后后迭代了五六个版本；三是对比实验法，把应用前后的批改时长、反馈周期、评价一致性、学生满意度这些量化数据做前后对比。

3 评价智能体的设计与构建

3.1 评价指标体系设计

评价指标体系是评价智能体运行的根本依据。本研究围绕中职数字媒体专业 Photoshop 课程的教学内容，对瑕疵处理、抠像换背景、色彩调整、材质合成、综合项目五种典型作品类型进行特征分析，提炼出各类型的共性评价

维度和专项评价要点，形成了分层次、分类型的指标框架，共计 15 个评价维度，每个维度下设优秀、良好、合格、待改进四个等级，各等级附有明确的特征描述（见表 1）。

表1 评价指标体系概览

作品类型	评价维度	维度数
瑕疵处理	污点消除彻底度、肤质保留自然度、整体画面协调性	3
抠像换背景	边缘处理精细度、细节保留程度、与新背景融合效果、投影与环境光匹配	4
色彩调整	色彩准确性、色彩层次感、整体色调和谐度、创意表现力	4
材质合成	材质感真实度、光影一致性、透视合理性、创意表达	4
综合项目	主题表达、技术综合运用、视觉美感、实用价值	4

各维度的等级特征描述力求准确、具体、可观察。以抠像类作品中“边缘处理精细度”维度为例：优秀等级要求“边缘干净利落，无可见锯齿或杂边，细节保留完整”，良好等级为“边缘基本整洁，仅在复杂区域存在少量不完美之处”，合格等级为“边缘存在明显锯齿或杂边，需要进一步优化”，待改进等级则为“边缘处理粗糙，存在大量锯齿、杂边或细节丢失”。这种结构化的等级定义方式，既保障了评价的全面性和一致性，也为智能体的理解和执行提供了清晰标准。

3.2 评价智能体的技术实现

本研究基于扣子平台构建评价智能体，技术实现围绕提示词工程、知识库配置和工作流编排三个核心环节展开。

3.2.1 提示词工程设计

提示词我前后改了不下 5 版。在最初比较抽象的版本时，智能体给出的反馈要么太空、要么抓不住重点。之后我把每条评价指标拆成“找问题—给建议—列教程”三个步骤，让智能体先指出学生作品具体差在哪里，再给出改进方向，最后附上对应的操作要点。这样，输出的报告才有了“对症下药”的感觉。

3.2.2 知识库配置

知识库用的是扣子自带的 RAG 能力，我先把 Photoshop 里常用的工具操作文档、我的教学案例、常见错误示范整理成知识文档传进去。举个例子：学生抠像的边缘总是处理不干净，智能体识别出该问题后，会自动从知识库调出钢笔工具的精细抠图步骤推给学生。等于智能体不光会挑错，还会把“怎么改”的方法传给学生。

3.2.3 工作流编排

工作流模拟了我批改作业的顺序。学生传作品图，视觉理解模型先判断是哪一类作品，匹配对应的评价指标，

再按预设的权重对各维度打分，合成综合分。最后把评分、问题清单、改进建议、操作教程打包成一份结构化报告推给学生，学生如果有疑问可以继续追问。

3.3 人机协同评价流程设计

整个评价流程我设计成一个闭环：AI 初评→教师复核→学生修改→AI 复评。第一轮 AI 初评，智能体把多维度分析和评分自动跑完，生成报告里包含具体问题清单和改进建议；教师复核时主要看两件事——一是 AI 在创意性、艺术性这类维度上判断不了的内容，二是结合企业或客户实际需求看作品是否达标^[3]。从整个设计看，AI 是来帮忙的，不是来替代的^[3]。

4 实践应用与效果分析

为验证评价智能体的实施效果，本人在所任教的两个班级开展了为期一个学期的课程应用实践，从批改时长、评价标准的一致性、反馈及时性满意度、学生作品质量等几个维度收集数据（见表 2）。

表2 评价智能体应用前后效果对比

评价指标	应用前	应用后	变化幅度
单班批改时长	约92分钟	约25-40分钟 (复核)	节省超50%
学生获反馈周期	约1周	当天	显著缩短
评价标准一致性系数	0.72	0.89	+0.17
反馈及时性满意度	58%	89%	+31%
作品及格率	87%	94%	+7%
作品优良率	31%	42%	+11%

4.1 批改效率显著提升

引入评价智能体之前，两个班级的批改工作完全依赖教师逐份审阅，单次作业批改平均耗时约 92 分钟，学生需等待一周左右才能收到反馈。引入之后，评价方式从教师单向批改转变为学生自主参与、AI 即时反馈与教师针对性指导的协同模式。学生提交作品后，智能体约 30 秒即可完成一份作品的初评，40 人班级的全部初评在 20 分钟左右即可完成，学生当天就能获取评价反馈并着手修改。教师的工作重心则从逐份批改转向复核抽检和个性化指导，得以将更多精力投入到创意表达指导和课堂互动之中。

4.2 评价标准一致性明显改善

为量化评价标准一致性的变化，本研究设计了标准一致性检验方案：随机抽取 20 份作品，由同一教师在不同时间点进行两次人工评分，计算两次评分的相关系数；同时由评价智能体对同一批作品评分，计算智能体评分与教师评分的一致性系数。数据显示，引入评价智能体之前，同一教师在不同时段的评分相关系数约为 0.72，存在一定程度的评分漂移；引入之后，教师在复核环节的评分与智能

体初评的一致性系数达到 0.89，跨时段一致性显著提升，评分漂移问题得到有效缓解^[2]。

4.3 反馈及时性与质量获得认可

反馈及时性方面，学生从提交作业到获得反馈的周期由原来的一周左右缩短至当天甚至数小时内，教师复核后的完整反馈通常可在 48 小时内返回。反馈周期的缩短使学生能够在学习记忆较为鲜明的时间窗口内进行反思和修改，有效促进了反馈对学习的驱动作用。反馈质量方面，智能体生成的反馈较传统人工评语更为具体。例如，针对边缘处理问题，智能体会指出“左侧头发边缘存在明显的背景残留，建议使用钢笔工具重新勾勒选区并适当羽化处理”，并可推荐知识库中的操作教程。学生认为智能体反馈“具体明确”“有参考价值”的比例达 82%，明显高于人工反馈的相应比例。

4.4 实践中的不足与反思

本研究在实施中也发现了一些待改进之处：其一，评价智能体对创意性、艺术性较强作品的评价能力仍有局限，在涉及审美判断的维度上仍需依赖教师的专业复核^[3]；其二，智能体生成的部分反馈建议存在重复化、模板化的倾向，需进一步优化提示词以提升个性化程度^[4]；其三，学生对智能体反馈的信任度需要一个培养过程，部分学生最初对 AI 评价结果持观望态度，需通过实际体验逐步建立信任。

5 结论与展望

经过一学期的实践检验，可以从四个维度对这项工作作出总结。其一，评价范式的转变。作品批改不再是教师的单向过程，而是学生提交、AI 即时初评与教师针对性复核三方协同的实践。学生身份也随之改变——从被动等待反馈的接受者，成为带着初评意见主动修改的参与者。反馈周期可从一周压缩到当天，这一变化的价值在于：学生能在记忆最为鲜明的时候完成反思与优化。其二，评价一致性得到改善，评分漂移得到了有效抑制。这一改善是人机协同过程中评价智能体与教师评分相互校准的结果。其三，人机分工的边界得到进一步清晰，技术性、规则性等维度由智能体来承担，教师则专注于创意判断、需求评估、个性化指导等主观性更强的工作，人工智能的效率优势和教师的专业判断形成了互补。其四，实施路径的可推广性得到了初步验证。整套方案不依赖编程能力，核心门槛在于将教学经验翻译为提示词和知识库文档，能在同类中职数字媒体课程中具有较好的迁移价值。

展望未来，评价智能体应结合学习轨迹数据实现更精

准的学习诊断。应用层面,可将其从单一课程延伸到课程群,逐步构建较为完整的数字媒体专业智能评价体系。同时评价中 AI 与教师的最佳分工比例等问题,也值得进一步深入探讨。

参考文献:

[1] 本书编写组. 职业教育专业目录(2021 年)[M]. 北京: 高等教育出版社, 2021:5.

[2] 臧铁军, 刘素杰. AI 时代教育评价发展趋向[J]. 基础教育课程, 2025(9):19-25.

[3] 王辞晓, 伍潇贝, 尚俊杰. 兼顾多元主体质量观的中小学人机协同教学质量评价框架研究[J]. 电化教育研究, 2025, 46(9):83-90,98.

[4] Qian Y. Prompt Engineering in Education: A Systematic Review of Approaches and Educational Applications[J]. Journal of Educational Computing Research, 2025, 63(7-8): 1782-1818. DOI:10.1177/07356331251365189.

[5] ElSayary A, Kuhail M A, Hojeij Z. Examining the Role of Prompt Engineering in Utilizing Generative AI Tools for Lesson Planning: Insights From Teachers' Experiences and Perceptions[J]. Human Behavior and Emerging Technologies, 2025: 9986139. DOI:10.1155/hbe2/9986139.

[6] Gao Y, Xiong Y, Gao X, et al. Retrieval-Augmented Generation for Large Language Models: A Survey[J]. arXiv preprint arXiv: 2312. 10997, 2023.