

图像生成评估：人类评估与自动评估的全面综述

刘绮¹，杨双林²，李泽健¹，侯乐凡³，孟辰焯³，张颖¹，孙凌云³

¹浙江大学软件学院，中国宁波市，315100

²东南大学计算机科学与工程学院，中国南京市，211189

³浙江大学计算机科学与技术学院，中国杭州市，310027

摘要：图像生成模型取得了显著进展，其中图像评估在解释和推动这些模型的发展方面至关重要。现有研究广泛探讨了图像生成的人类评估与自动评估。本文对相关研究进行了系统综述，重点涵盖两个核心部分：评估协议与评估方法。首先，总结了10类图像生成任务，重点关注它们在评估方面的差异。基于此，提出一种新的评估协议，以涵盖不同图像生成任务所需的人类与自动评估的重要评估方面。其次，重点回顾过去5年中提出的自动评估方法。据我们所知，本文是对人工评估的首次全面总结，涵盖评估方法、工具、评估细节及数据分析方法。最后，探讨了当前图像生成评估面临的挑战及未来发展方向。希望本综述能够帮助研究人员系统理解图像生成评估，掌握该领域最新进展，并推动相关研究的开展。

关键词：图像生成评估；人类评估；自动评估；评估协议；评估方面
<https://doi.org/10.1631/FITEE.2400904>