意识形态的回声——探索历史纳粹宣传电影中人物特质的音频分析流程

Nicolas Ruth[1],Manuel Burghardt[1]

[1] 莱比锡大学计算人文组，德国莱比锡

摘要

本研究探讨如何运用计算音频分析考察纳粹宣传电影中的意识形态叙事。研究采用三步流程——说话人分离、音频转录和心理语言学分析——揭示影片人物中的意识形态模式。尽管目前说话人分离仍存在问题，该方法仍能深入揭示人物特质与宣传叙事，并显示出可扩展应用的潜力。

关键词： 数字电影研究、人物分析、数字历史、计算音频分析

arXiv:2601.08879v1 [cs.SD] 12 Jan 2026

1 引言

纳粹政权将电影用作宣传工具，充分体现了其对电影艺术无与伦比的力量的信念：电影能够塑造思想、传播意识形态。纳粹分子践行约瑟夫·戈培尔提出的“思想革命”理念（Leiser，转引自 [15]），将所有艺术与文化纳入其世界观的支配之下，并建立起集中且受到严密控制的电影产业，以制作具有意识形态导向的电影，甚至倡导一种全新的电影风格 [13]。如今，仍有44部影片被正式列为禁映纳粹宣传电影（限制放映影片），原因是它们具有特别重大的意识形态与历史意义。这些影片由弗里德里希·威廉·茂瑙基金会保存，仅可依据严格的学术研究协议观看。有趣的是，同一时期的许多其他影片，例如德国经典影片 Die Feuerzangenbowle（火钳酒）（英文名：火钳酒），至今仍在每年圣诞节播出，且广泛可见。我们并没有明确的标准来判断哪些纳粹电影应被归为无害，哪些不应如此归类。相反，战后德国电影审查的整个历史深受社会变迁和各种不同影响的塑造，例如来自不同占领区的影响[17]。关于宣传电影与非宣传电影之间的区别，至今仍存在争论。近年来，历史学界也重新关注纳粹时期电影的遗产。2025年初，德国一家领先的历史研究机构启动了一个名为“Screens”的新项目，将考察这些影片的各种元素及其意识形态基础。[1] 本研究旨在分析若干部限制放映的禁映纳粹宣传电影，以了解它们如何构建意识形态叙事。研究对象包括《犹太人苏斯》（1940）（英文名：《犹太人聚聚》）、Hitlerjunge Quex（希特勒青年队员克维克斯）（1933）（英文名：《希特勒青年队员克维克斯》）以及 Kopf hoch, Johannes!（振作起来，约翰内斯！）（1941）（英文名：《振作起来，约翰内斯！》）。我们将采用计算分析方法，揭示这些影片中的意识形态模式，并最终把这一方法扩展到数百部公开可看的纳粹时期电影，以探究其中实际包含了多少“隐性”意识形态。

Nicolas Ruth 和 Manuel Burghardt。《意识形态的回声——构建音频分析流程，以揭示历史纳粹宣传电影中的人物特质》。载于：*ADHO* 数字人文会议论文集（*DH2025*），*2025* 年，葡萄牙里斯本。© 2025 年，作者版权所有。依据知识共享署名 4.0 国际许可协议（CC BY 4.0）授权。[1]

https://www.ifz-muenchen.de/aktuelles/artikel/screens-spielfilme-als-zeithistorische-quelle

2 方法论考量

在电影研究领域，人们普遍认为，电影是一种多模态 [2] 且高度复杂的现象，兼具视觉层与音频层；后者又可进一步分为口语、音乐和音效。我们认为，数字电影研究已对视觉层进行了广泛研究，并由此形成了“远观” [1] 和“深度观看” [3],等概念。这些方法运用现代深度学习技术，推动了所谓的“数字视觉转向” [26]。然而，我们认为，在音频计算分析方面，尤其是历史电影的相关研究中，仍存在显著的方法论空白。因此，我们提出并评估一个专门用于历史纳粹电影音频分析的三步分析工作流程。目前，我们聚焦于人物对话，因为我们认为，意识形态模式主要通过电影人物及其口语互动传达。对纳粹时期电影中电影人物塑造方式的研究，带来了诸多研究机会。其中两个关键领域包括：分析对按族裔划分的内群体与外群体的文化宣传，其中（例如反犹主义）人物描绘会赋予其歧视性特征 [22], ；以及分析女性气质的呈现方式，在纳粹德国，这被视为文化身份的体现 [12]。通过将人物描绘与人工选定的参照对象进行自动比较，并在音频分析流程中应用语音性别检测，可以获得相关见解。我们对人物特质进行的计算分析，在方法论上与“心理语言学”领域相关。 [16] 描述了计算机辅助自然语言处理方法日益兴起的应用，即以定量方式评估基于心理学的人物特质，从而弥合定性研究与定量研究之间的差距。这些方法包括主题建模等技术，可用于传统心理学访谈；也包括 LIWC [7] 和 Empath [9], 等基于词典的方法，通过语言元素推断特质。一些研究者甚至认为，语言可能比传统的心理自我报告更能体现人格 [5, 6]。在大规模语言语料库的计算分析方面，近期基于 Transformer 的语言模型取得的进展表明，在情感分析 [4]等任务中，这类模型的表现始终优于传统的基于词典的方法。基于这些研究，我们采用基于大语言模型的方法，分析电影人物的人物特质。尽管虚构人物使用的是剧本语言，但这恰恰是我们关注的重点，因为我们要探究国家社会主义如何塑造人物描绘。

3 音频分析工作流程

我们建议采用三步流程（见图 1），将影片中的口语对白自动转换为人物分析，并评估每个步骤中最先进的深度学习模型。

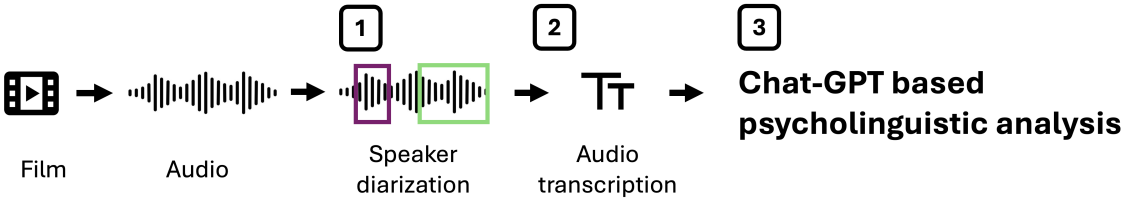

图 1：三步分析工作流程概览。

首先，从影片中提取音频层，并将其送入专门用于（1）说话人日记的步骤，即识别不同说话人的过程。该过程包括语音活动检测（VAD）、计算说话人嵌入，以及根据说话人身份对语音片段进行聚类。这项任务尤其具有挑战性，因为我们处理的是1933年至

1945年。为进行分析，我们评估了两种广受欢迎的开源工具：Pyannote 的[2] 说话人日记工具和 Nvidia Nemo[3] [10]。流程的下一步是（2）音频转录，即使用自动语音识别（ASR）将音频转换为与已识别说话人相关联的文本。我们评估了 OpenAI 的 Whisper [21] 的四个版本：Whisper-medium、Whisper large-v2，以及 Bofeng Huang 的德语微调版本[4] 。Whisper 被认为是最先进的模型，以其对背景噪声 [18]的鲁棒性而著称。最后，使用基于GPT的方法，对每位说话人收集到的讲话进行（3）“心理语言学”人物特质分析。此过程将每位（主要）说话人的语音片段汇总成列表，并通过应用程序接口提交给 gpt-3.5-turbo-0125，使用以下提示词（由原始德语提示词翻译成英文）：

System prompt:
“You are an expert in analyzing fictional characters. Analyze this character monologue and answer these questions if possible. Keep it short and give your answer in the format of a Python dictionary with the numbering as keys!

1. Describe the character's personality in five words.
2. What is the most important character trait? Answer in one word.
3. What are the character's main goals?
4. What are the character's main motivations?
5. How does the character interact with other characters?
6. Name the title of the film if you recognize it!

Text of the character:”

Input: <INSERT TRANSCRIBED SPEECH SEGMENTS AS LIST>

图 2： 基于 GPT 的人物分析所用的系统提示词和输入。

出于应用程序接口成本方面的考虑，我们选择了较小的模型（gpt-3.5-turbo-0125）。通过为各段发言（ >2段）创建发言频率直方图，并选取频率高于上围栏的发言者来确定主要发言者，从而确保方法能够适应影片的发言模式，并为心理语言学分析提供充足的材料。

4 方法评估

4.1 步骤 1：说话人分离

虽然说话人识别系统通常是为相对受控的环境而设计的，但本研究面临着更多复杂性：历史德语音频、带有背景噪声的电影音频、不确定的说话人数，以及过长的音频信号。此外，复杂框架中的模型及其可单独配置的步骤还带来了更多方法论上的困难和潜在错误。基于音频的说话人识别系统通过Pyannote的多种配置进行了测试，包括其标准版本、两种语音隔离技术[5], 、德语分段模型[6], 、在《犹太人苏斯》上微调的分段模型，以及一种通过人脸嵌入聚类纠正说话人混淆的定制方法。此外，还对Nemo模型（通用配置、固定30名说话人）进行了评估，其表现优于所有Pyannote配置，平均说话人日记错误率

[2] https://github.com/pyannote/pyannote-audio[3] https://docs.nvidia.com/nemo-framework/user-guide/latest/nemotoolkit/asr/speaker_diarization/intro.html[4] https://huggingface.co/bofenghuang/whisper-medium-cv11-german & https://huggingface.co/bofenghuang/whisper-large-v2-cv11-german[5] https://pypi.org/project/noisereduce/ & https://huggingface.co/speechbrain/sepformer-dns4-16k-enhancement[6] https://huggingface.co/diarizers-community/speaker-segmentation-fine-tuned-callhome-deu

（DER）为58.84。DER通过错误归属的语音时长占比来衡量说话人分离性能，其中包括错误检测、漏检语音或错误聚类；由于存在重叠语音，其数值可能超过100% [11]。迈向完全自动化的这一步是未来需要解决的最大挑战，因此我们工作流程中的后续步骤依赖于为说话人识别手动创建的真实标注。对于这项任务，针对复杂视频材料存在多种方法，包括需要使用这些数据进行测试的跨模态模型 [8, 23]。

4.2 步骤2：转录

关于自动语音识别，这些模型对背景噪声表现出很强的鲁棒性，并且能够准确捕捉标点符号，从而有助于理解句子的含义。主要错误源于较为古老的德语词汇（例如Durchlaucht、Staatsstreich、Rabbuni）、专有名词、方言/语言（例如意第绪语、柏林方言），以及会引发自动语音识别幻觉的过于轻声的语音。经过德语微调的模型在处理 Jude、Jid 和 Jiden 等特定德语词汇时表现略好。不过，总体而言，两个大模型版本都很有效，能够得到良好应用。以下对基于GPT的分析进行的评估表明，尽管存在少量错误，转录结果仍足以满足预期流程的要求。研究（例如 [25], ）表明，自动语音识别错误和排版错误可能会降低大语言模型的性能。然而，我们的应用似乎能够抵御此类问题，因为自动语音识别错误率足够低。

4.3 步骤3：ChatGPT分析

基于GPT的分析首先揭示，该模型无法根据提供的人物台词识别影片；这一点已通过提示中的第6个问题进行测试（见图2）。唯一的例外是台词中直接提到了一位相当著名的主角，而被提及的人不一定就是说出这句台词的角色。《Hitlerjunge Quex》中 Heini Völker 和《犹太人苏斯》中的 Joseph Süß Oppenheimer 就属于这种情况。除此之外，模型会凭空编造随机的电影名称，这表明分析结果并非基于对这些影片的先验知识。

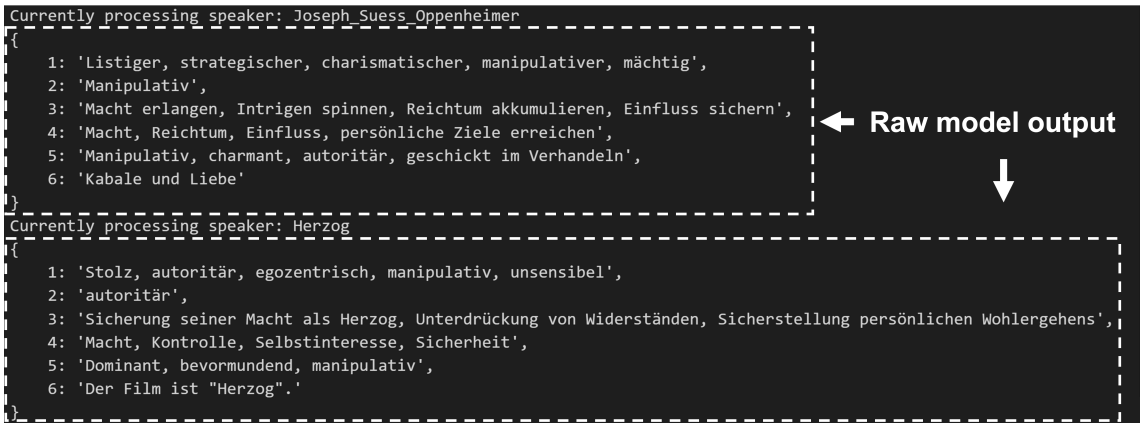

图3：针对“Joseph Süß Oppenheimer”和公爵这两个角色的基于GPT的人物分析输出，后译为英文以供论文使用。

图3展示了反犹宣传电影《犹太人苏斯》中两个角色的直接分析结果。该方法证明了其分析角色的准确性，仅凭语音片段便能深入

洞察他们的人格。对约瑟夫·聚斯·奥本海默的分析将他描绘为一个狡猾、善于谋略、富有魅力且权势强大的角色，他通过操纵、蛊惑他人并以权威姿态与人交往来获取权力、财富和影响力。这些分析结果与有关这部电影的学术文献相符；文献指出，聚斯·奥本海默的形象被塑造成一种漫画式的反犹人格 [20]。对公爵的刻画同样可以认为是准确的：他骄傲、专横且自我中心，为维护自身权力和个人利益而表现得盛气凌人、居高临下且善于操纵。另一个令人印象深刻的例子来自另一部电影：纳粹青年电影《Kopf hoch, Johannes!（振作起来，约翰内斯！）》中的朱丽叶塔·默克、约翰内斯·冯·雷德尔及约翰内斯的父亲构成的三角关系。相关文献这样描述主要人物之间的关系：约翰内斯·冯·雷德尔在母亲去世后回到德国，难以适应环境，并与粗暴、情感疏离的父亲关系冷淡。他的姑妈朱丽叶塔·默克试图弥合父子之间的隔阂 [13]。文献中的人物三角关系已在图4中以人工方式呈现，并结合自动分析得出的、经翻译的人格特质、目标和互动结果进行了补充，展示了该方法捕捉复杂人物特质的能力。

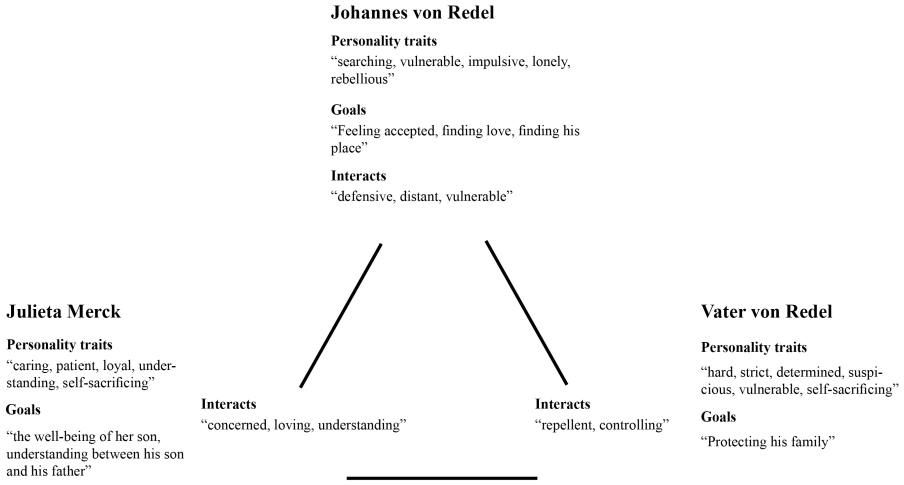

图4：纳粹青年电影《Kopf hoch, Johannes!（振作起来，约翰内斯！）》中朱丽叶塔·默克、约翰内斯·冯·雷德尔与约翰内斯父亲之间的人物关系三角。

5 结论

这表明，针对德国历史电影素材进行音频分析的拟议工作流程，在说话人分离方面仍需进一步完善，但在音频转录和人物分析任务上已表现出色。关于说话人识别的未来发展，除了前文提到的多模态方法外，还可以更深入地关注音频数据，基于历史音频样本开展专门的微调和有针对性的模型开发。当前音频深度学习模型的发展带来了令人振奋的机遇，尤其是若能将 Whisper 作为主流自动语音识别模型的优势应用于其他音频任务。音频转录的表现稳健，准确度也足够高。影片中关键人物特质的识别结果，在我们的测试中与相关文献中的既有分析大体一致，这为定量分析纳粹时期电影中的角色发展提供了可能，也有助于深入了解纳粹意识形态；这类分析还可扩展至该时期更大规模的电影语料库，我们认为其中隐藏着更多意识形态。对于大规模的人物特质分析，我们将探索使用 GPT 的封闭词汇表或词嵌入系统。最终，我们希望更全面地审视纳粹电影中的意识形态，并计划

同时采用最先进的方法，对视觉层进行远观，正如当前工具 Vian [14], Tibava [24] or Zoetrope [19]所展示的那样。

伦理声明

本研究涉及对禁映纳粹宣传电影（受限制的纳粹宣传电影）的分析，其中包含反犹、种族主义及其他有害内容。这些影片由弗里德里希·威廉·茂瑙基金会保存，仅在受控条件下供学术研究使用。本文的所有分析均在批判性学术框架内进行，旨在理解历史电影艺术中的意识形态建构机制。本文未公开传播任何源自这些影片的材料，也未试图在本研究范围之外复制或传播有害内容。本研究遵循负责任的数据处理和尊重历史的伦理原则，同时认识到这些材料的持续影响及其可能遭到滥用的风险。

致谢

我们感谢弗里德里希·威廉·茂瑙基金会授权我们使用本研究所需的数字化禁映纳粹宣传电影。基金会的支持对于拟议音频分析流程的开发与评估至关重要。

参考文献

[1] Arnold, Taylor 和 Tilton, Lauren。《远观：分析大型视觉语料库》。载于：《数字人文研究》第*34*卷，增刊_1 （2019年），第 i3–i16 页。[2] Bateman, John 和 Schmidt, Karl-Heinrich。《多模态电影分析：电影如何表达意义》。纽约：劳特利奇出版社，2013年。[3] Bermeitinger, Bernhard 等。《深度观看：探索文化研究中分析视觉媒体的新方法》。载于：《摘要集——乌得勒支 *ADHO*数字人文会议》。2019年。[4] Borst, Janos、Klähn, Jannis 和 Burghardt, Manuel。《词典消亡？——零样本情感分类的兴起》。载于：《*2023*年计算人文研究会议论文集》。第3558卷。CEUR研讨会论文集。2023年，第303–319页。[5] Boyd, Ryan L 和 Pennebaker, James W。《基于语言的人格：数字世界中人格研究的新方法》。载于：《行为科学最新观点》第*18*卷 （2017年）。行为科学中的大数据，第63–68页。ISSN：2352-1546。DOI：https://doi. org/10.1016/j.cobeha.2017.07.017。URL：https://www.sciencedirect.com/ science/article/pii/S2352154617300487。[6] Boyd, Ryan L.、Pasca, Paola 和 Lanning, Kevin。《人格全景：通过大型行为数据对人格进行概念化》。载于：《欧洲人格杂志》第*34*卷，第5期（2020年），第599–612页。DOI：10 . 1002 /per . 2254。URL：https : / / doi . org/10 . 1002/per.2254。[7] Boyd, Ryan L. 等。《LIWC-22 的开发与心理测量特性》。技术报告。访问日期：2025年10月17日。德克萨斯大学奥斯汀分校，2022年。URL：https://www.liwc.app。

[8] Cheng, Luyao 等。《整合音频、视觉和语义信息，提升多方对话中的多模态说话人分离》。载于：《计算语言学协会第*63*届年会论文集（第一卷：长篇论文）》，Wanxiang Che 等编。奥地利维也纳：计算语言学协会，2025年7月，第19914–19928页。ISBN：979-8-89176-251-0。DOI：10.18653/v1/2025.acl- long.977。URL：https://aclanthology.org/2025.acl-long.977/。[9] Fast, Ethan、Chen, Binbin 和 Bernstein, Michael S.《Empath：理解大规模文本中的主题信号》。载于：《*2016*年*CHI*计算机系统人因会议论文集》。CHI ’16。美国加利福尼亚州圣何塞：计算机协会，2016年，第4647–4657页。ISBN：9781450333627。DOI：10.1145/2858036.2858535。URL：https://doi.org/10.1145/2858036.2858535。[10] Fischbach, Lea。《说话人分离模型的比较分析：为德语方言语音创建数据集》。载于：《第三届自然语言处理在田野语言学中的应用研讨会论文集》，Oleg Serikov 等编。泰国曼谷：计算语言学协会，2024年8月，第43–51页。DOI：10.18653/v1/2024.fieldmatters-1.6。URL：https://aclanthology.org/2024.fieldmatters-1.6/。[11] Fiscus, Jonathan G. 等。《2005年春季会议丰富转录识别评测》。载于：《多模态交互的机器学习》，Steve Renals 和 Samy Bengio 编。柏林、海德堡：Springer Berlin Heidelberg，2006年，第369–389页。ISBN：978-3- 540-32550-5。[12] Frietsch, Elke。《“人民与领袖”的媒介呈现：声学——图像——雕塑》。载于：《国家社会主义与性别》，Elke Frietsch 和 Christina Herkommer 编。比勒费尔德：transcript Verlag，2015年，第199–222页。[13] Giesen, Rolf 和 Hobsch, Manfred。《希特勒青年队员克维克斯、《犹太人苏斯》与《科尔贝格》：第三帝国宣传电影——纳粹电影文献与资料》。柏林：Schwarzkopf & Schwarzkopf Verlag，2005年。[14] Halter, Gaudenz 等。《VIAN：用于电影分析的视觉标注工具》。载于：《计算机图形论坛》。第38卷第3期。Wiley Online Library。2019年，第119–129页。[15] Hardinghaus, Christian。《电影宣传大屠杀？——以煽动性电影《永恒的犹太人》和《犹太人苏斯》为例的研究》。马尔堡：Tectum-Verl.，2008年。[16] Herderich, Alina、Freudenthaler, Heribert H. 和 Garcia, David。《一种从文本数据中揭示心理构念的计算方法》。载于：《心理学方法》（2024年）。ISSN：1082-989X。DOI：10.1037/met0000700。[17] Hoppe, Johanna。《大炮与麻雀：1945年后遭禁纳粹电影的论述史》。马尔堡：Schüren Verlag，2021年。[18] Kuhn, Korbinian 等。《衡量自动语音识别解决方案的准确性》。载于：第*16*卷，第4期（2024年1月）。ISSN：1936-7228。DOI：10.1145/3636513。URL：https: //doi.org/10.1145/3636513。[19] Liebl, Bernhardt 和 Burghardt, Manuel。《Zoetrope——新闻视频中的交互式特征探索》。载于：《数字人文*2023*：摘要集》，Walter Scholger 等编。日内瓦：计算语言学协会，2024年，第432–434页。DOI：https: //doi.org/10.5281/zenodo.8107770。[20] Niven, Bill。《犹太人苏斯：一部宣传电影的漫长生命》。哈勒：Mitteldeutscher Verlag，2022年。

[21] Radford, Alec 等. “通过大规模弱监督实现鲁棒语音识别。”载于：第*40*届国际机器学习会议论文集。ICML’23。美国夏威夷州檀香山：JMLR.org，2023年。[22] Schwarz-Friesel, Monika 和 Reinharz, Jehuda。*21*世纪仇犹主义话语。柏林：德格鲁伊特出版社，2017年。[23] Sharma, Rahul、Somandepalli, Krishna 和 Narayanan, Shrikanth。“用于弱监督主动说话人定位的跨模态视频表征。”载于：《*IEEE*多媒体汇刊》第*25*卷 (2023年)，第7825–7836页。DOI：10.1109/TMM.2022.3229975。[24] Springstein, Matthias 等。“TIB AV-Analytics：用于学术视频分析和电影研究的基于网络的平台。”载于：第*46*届国际*ACM SIGIR*信息检索研究与开发会议论文集。2023年，第3195–3199页。[25] Wang, Bin 等。“大语言模型对噪声指令的鲁棒性。”载于：计算语言学协会会议论文集：*EMNLP 2024*，由 Yaser Al-Onaizan、Mohit Bansal 和 Yun-Nung Chen 编。美国佛罗里达州迈阿密：计算语言学协会，2024年11月，第11939–11950页。DOI：10.18653/v1/2024.findings-emnlp. 697。URL：https://aclanthology.org/2024.findings-emnlp.697/。[26] Wevers, Melvin 和 Smits, Thomas。“视觉数字化转向：利用神经网络研究历史图像。”载于：《数字人文研究》第*35*卷，第1期（2019年1月），第194–207页。ISSN：2055-7671。DOI：10.1093/llc/fqy085。电子版：https://academic.oup.com/dsh/article-pdf/35/1/194/32976784/fqy085.pdf。URL：https://doi.org/10.1093/llc/fqy085。