arXiv:2304.14495v1 [cs.CV] 3 Apr 2023

生理与医疗保健神经网络中的模型可解释性

Rohit Sharma[1],Abhinav Gupta[2], Arnav Gupta[2], 和 Bo Li[2]

[1] 肯塔基大学，美国[2] 腾讯，美国
{boli, agupta}@tencent.com

摘要。 SpO2估计与监测对于评估肺功能和治疗慢性肺部疾病至关重要。COVID-19疫情凸显了及早发现SpO2变化的重要性，尤其是对于病情出现临床恶化的无症状患者。然而，传统的血氧饱和度测量方法依赖接触式传感，存在交叉污染风险，并可能给肢体灌注受损的患者带来并发症。此外，边缘化社区和欠发达国家可能无法获得脉搏血氧仪。为克服这些局限，并以更舒适、不易察觉的方式监测SpO2，近期研究探索了利用视频测量SpO2。然而，使用摄像头以非接触方式测量SpO2，尤其是通过智能手机进行测量，颇具挑战，因为智能手机摄像头传感器产生的生理信号较弱，光学选择性较低。该系统包含三个主要步骤：1）从智能手机拍摄的视频中提取感兴趣区域（ROI），包括手掌和手背；2）对ROI进行空间平均，生成R、G和B时间序列；3）将时间序列输入光学生理学启发的卷积神经网络，以进行SpO2估计。我们提出的方法能够利用消费级智能手机拍摄的视频，更高效、准确地监测SpO2，在远程医疗和健康筛查场景中尤其适用。

关键词： 机器学习、远程光电容积描记法、医疗保健人工智能

1 引言

血氧饱和度的估算和监测对于评估肺功能和治疗慢性肺部疾病至关重要。COVID-19疫情凸显了及早发现血氧饱和度变化的重要性，尤其是对于临床恶化的无症状患者。然而，传统的血氧饱和度测量方法依赖接触式传感，存在交叉污染风险，并可能给肢体灌注受损的患者带来并发症。此外，边缘化社区和欠发达国家可能无法获得脉搏血氧仪。

为了解决这些局限，并以更舒适、无干扰的方式监测血氧饱和度，近期研究开始探索利用

视频测量血氧饱和度。然而，以无接触方式利用摄像头测量血氧饱和度，尤其是使用智能手机时，由于生理信号较弱，以及智能手机摄像头传感器的光学选择性较低，因而具有挑战性。以往的方法依赖于显式特征提取，作为脉搏血氧仪原理的一种变体。

本研究提出了一种创新方法，利用可解释的卷积神经网络（CNN）模型测量血氧饱和度。我们的方法从使用消费级智能手机摄像头拍摄的非接触式视频的全部三个颜色通道中进行整体特征提取。借助卷积神经网络模型，我们能够应对非接触式血氧饱和度测量的挑战，提高血氧饱和度监测的准确性和效率。这种方法有望应用于健康筛查和远程医疗，尤其是在传统接触式方法和脉搏血氧仪并不普及的地区。

近年来，人们越来越关注以无创且不易察觉的方式利用视频监测血氧饱和度，并将其应用于健康筛查和远程医疗。然而，使用摄像头进行非接触式血氧饱和度测量，尤其是使用智能手机摄像头时，会面临诸多挑战，因为生理信号较弱，且智能手机摄像头传感器的光学选择性较低。大多数既往研究采用显式特征提取方法，作为脉搏血氧仪原理的一种变体。与之不同，本文提出了可解释的卷积神经网络（CNN）模型，从使用消费级智能手机摄像头拍摄的非接触式视频的全部三个颜色通道中整体提取特征，以测量血氧饱和度。所提出的方法旨在克服以往方法的局限性，并提供一种更准确、高效的视频血氧饱和度监测方式。

图1展示了所提出的利用智能手机视频和卷积神经网络进行非接触式SpO2监测系统。该系统包括三个主要步骤：1）从智能手机拍摄的视频中提取感兴趣区域（ROI），其中包括手掌和手背；2）对ROI进行空间平均，生成R、G和B时间序列；3）将时间序列输入光学生理学启发的卷积神经网络，以估计SpO2。所提出的卷积神经网络依据应用于生理传感的光与组织相互作用原理设计，从而增强了模型的可解释性。本研究将手部区域作为概念验证，因为与使用面部测量SpO2（这是以往研究中的常见方法）相比，使用手部引发的隐私顾虑较少。此外，在新冠肺炎疫情期间，按照佩戴口罩的指南，录制手部视频是进行健康状况筛查和数据收集的更安全方式。本文后文将概述本研究的贡献。

1. 我们采用数据驱动方法，通过可视化 RGB通道组合的权重来展示模型的可解释性。我们发现，神经网络学到的颜色波段选择与光学生理学方法中建议使用的颜色波段一致。2. 我们分析手部两侧和不同肤色对 SpO2估计质量的影响。研究结果表明，肤色可以

影响 SpO2估计的性能，而且手背可能比手掌提供更准确的结果。

3. 我们将光生理学启发的神经网络结构与专为此问题设计的最先进的神经网络结构进行比较，并实现了更准确的SpO2预测。我们提出的方法利用消费级智能手机拍摄的视频监测SpO2，能够提供更高效、更准确的方式，在远程医疗和健康筛查场景中尤其有用。

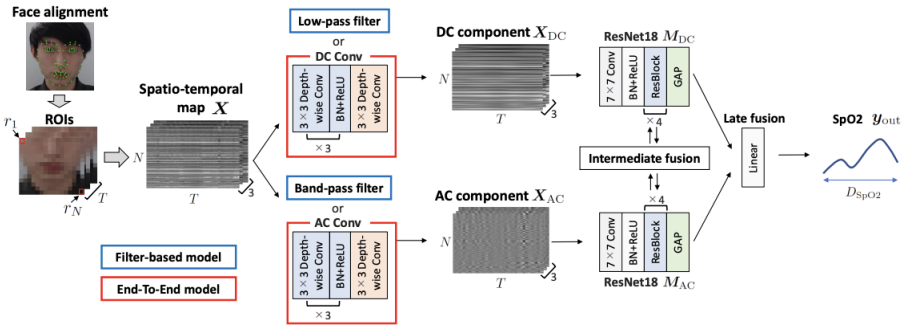

图1. 所提出的方法包括构建时空图，并通过滤波过程或卷积层提取直流分量和交流分量。

人体的血红蛋白（Hb）蛋白将氧气从肺输送到身体组织。血氧饱和度（SpO2）是血液中氧合血红蛋白（HbO2）占总血红蛋白的比例，用于指示呼吸系统的功能状况 [1]。SpO2的正常范围为95%至100% [1]。SpO2是衡量身体满足代谢需求能力的重要指标，SpO2下降可能意味着氧合不足和临床恶化 [1]。脉搏血氧测定是一种便捷、无创且可连续测量SpO2的方法 [4]。

脉搏血氧仪基于比值之比原理工作，该原理由 Aoyagi 于 20 世纪 70 年代首次提出 [4]。这些设备通过指尖发射两个波长的光（通常为红光和红外光），光线会与血液和组织相互作用，并被其衰减。光学传感器接收透射光，从中获取搏动性血容量信息。对两个波长下的搏动性血容量进行处理，以估算 SpO2。具体而言，SpO2 大致与比值之比成正比；该比值通过计算红光和红外光波长下透射脉搏信号的搏动性/交流（AC）分量与相对稳定的直流（DC）分量之比得出 [4]。脉搏血氧仪常用于医院、诊所和家庭。

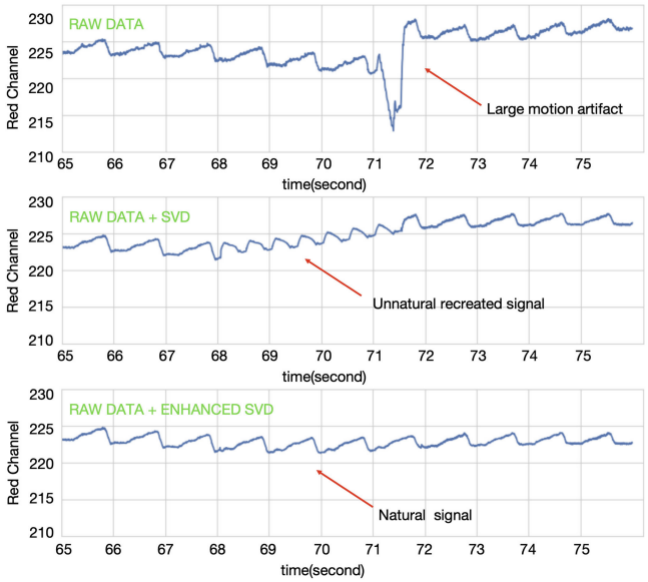

图 2. 不同伪影方法的信号比较。所提出的方法位于底部。

血氧饱和度（SpO2）是呼吸功能的重要指标，表示血液中氧合血红蛋白占总血红蛋白的比例。SpO2 异常下降可能表明氧合不足和临床状况恶化，因此它是需要重点监测的关键参数。脉搏血氧仪采用比率之比法，常用于无创、连续地测量 SpO2，但接触式测量可能令人不适，也可能引发卫生问题。因此，研究人员探索了非接触式基于视频的 SpO2 估算方法，以缓解这些问题 [10-30]。多年来，深度神经网络已被广泛用于从面部视频中远程获取心率 [5-10, 27-33]。

现有的非接触式方法可根据摄像头和光源的设置分为两类。第一类采用类似传统脉搏血氧测定的单色传感，使用高端单色摄像头或受控的单色光源。这些方法能够实现精确的血氧饱和度测量，但需要专用设备。第二类采用消费级 RGB 相机，例如数字网络摄像头和智能手机摄像头，这些设备更易于获取，但由于每个 R、G 和 B 颜色通道所感测的波长范围较宽，因此面临挑战。以往使用 RGB 相机的非接触式血氧饱和度测量方法主要依赖显式的特征提取和线性回归，而我们提出了一种利用神经网络从全部三个颜色通道中提取血氧饱和度信息的新方法。我们的方法受到用于血氧饱和度的光生理模型的启发

通过测量并运用深度学习，使用普通 RGB 摄像头以非接触方式监测血氧饱和度。与以往的非接触式方法相比，我们的方法具有多项优势，包括更高的准确性、更少的不适感和卫生问题，以及更广泛的适用性。我们通过可视化 RGB 通道组合的权重来展示模型的可解释性；这些权重也与光学生理学方法中建议使用的颜色波段一致。我们还分析了不同肤色和手部不同侧面对 SpO2估计的影响。

尽管基于视频的血氧饱和度监测深度学习仍处于起步阶段，但我们的方法展现出良好前景，有望应用于健康筛查和远程医疗。它提供了一种便捷、无创的方式，可持续监测血氧饱和度，并作为氧合不足和临床恶化的早期预警信号。因此，它有望改善患者预后并降低医疗保健成本。

2 方法

为了准确估算血氧饱和度，提取人体皮肤的颜色信息至关重要，因为与血氧饱和度相关的生理信息蕴含在皮肤反射/再发射光中。该过程包括从皮肤区域提取 R、G 和 B 时间序列，这些序列称为皮肤颜色信号。使用大津法将皮肤像素与背景区分开来，并确定感兴趣区域。通过对每个视频帧中的皮肤像素值取平均，生成 R、G 和 B 时间序列。随后，使用步长为 0.2 秒的滑动窗口将皮肤颜色信号划分为 10 秒片段，并将这些片段作为神经网络的输入。较长的片段有助于增强对传感噪声的抵抗能力。由于片段长度远超包含一个心跳周期血氧饱和度信息所需的最短长度，因此，全连接结构或卷积结构足以捕捉时间依赖关系，无需采用循环神经网络结构。为了构建易于理解和解释的血氧饱和度预测神经网络，我们提出采用逐层相关性传播（LRP） [8]。逐层相关性传播是一种常用技术，通过为输入特征分配相关性分数来解释深度神经网络的预测结果。该算法根据每个输入特征对最终输出预测的贡献，为其分配相关性分数，从而更好地理解哪些特征对预测最为重要。

在本研究中，我们将使用逐层相关性传播（LRP）来识别神经网络用于SpO2预测的相关特征。这将帮助我们理解对于准确的SpO2估计而言最为重要的生理信号。此外，通过分析相关特征，我们可以识别模型中潜在的噪声或偏差来源，并提高模型整体的鲁棒性和准确性。

总之，模型可解释性和可解释性对于医疗保健应用中 SpO2预测系统的设计至关重要。通过使用逐层相关性传播来识别神经网络用于预测的相关特征，我们可以

加深我们对底层生理信号的理解，并确保该模型值得信赖且具有临床相关性。

3 实验结果

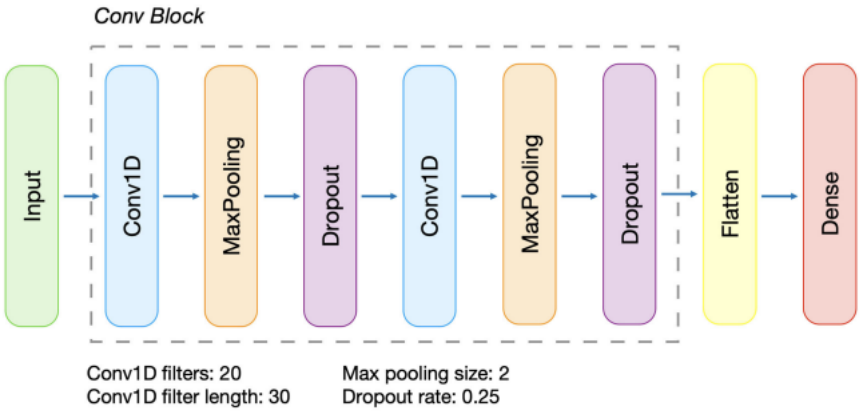

图 3. 使用网格搜索为一维卷积神经网络搜索参数。

其中一段记录用于模型的训练和验证，另一段记录则用于测试。为确保训练集和验证集之间不存在重叠的数据片段，每段记录被划分为三个呼吸周期，前两个周期用于训练，第三个周期用于验证。为准确估算血氧饱和度，从人体皮肤提取颜色信息至关重要，因为与血氧饱和度相关的生理信息蕴含在皮肤反射/再发射的光中。该过程包括从皮肤区域提取 R、G 和 B 时间序列，这些序列称为皮肤颜色信号。使用大津法从背景中区分皮肤像素，并确定感兴趣区域。对每个视频帧中的皮肤像素值取平均，生成 R、G 和 B 时间序列。随后，使用步长为 0.2 秒的滑动窗口，将皮肤颜色信号划分为 10 秒的片段，作为神经网络的输入。为解决参与者特定实验中数据集规模较小的问题，采用自助法数据复用策略，通过有放回抽样扩充训练和验证数据。这种过采样技术也有助于解决血氧饱和度数据值分布不均衡的问题。利用训练和验证数据调整模型结构及超参数后，使用调优得到的最佳超参数训练多个模型实例。在不同实例之间改变模型权重初始化和随机过采样所用的随机种子，以确保对模型进行全面评估。

随后，选择验证集均方根误差最高的模型实例，在测试记录上进行评估。我们采用了多种预处理技术，为每个颜色通道获取原始光电容积脉搏波信号、带通滤波信号和偏置信号。为自动从这些信号流中提取特征，我们使用了一维卷积神经网络（CNN）模型，该模型由两个时间卷积层组成，并采用最大池化和 25% 的丢弃率。我们采用交叉验证网格搜索来确定最佳超参数组合，其中包括卷积层数量、输入窗口大小、滤波器数量和滤波器长度。该卷积神经网络模型使用 Python 3.5 和 TensorFlow 1.2 实现，并在配备 NVIDIA P100 GPU 的 36 节点集群上进行训练；整个网格搜索过程耗时约 4 周。

参考文献

1. J. S. Zdanowicz，《通过数据挖掘检测洗钱和恐怖主义融资》，《ACM通讯》，第47卷，第5期，第53–55页，2004年5月。2. Johnson, E., Tang, X. 和 Samudrala, S., 2022. WeightMom：使用基于迭代动量的剪枝学习稀疏网络。arXiv预印本 arXiv:2208.05970。3. H. M. Alzoubi、M. In’airat 和 G. Ahmed，《探讨全面质量管理实践和六西格玛流程对提升质量、降低质量成本的影响：以迪拜为例》，《国际商业卓越期刊》，第27卷，第1期，第94–109页，2022年。4. B. Kurdi、M. Alshurideh、I. Akour、E. Tariq、A. AlHamad 和 H. Alzoubi，《社交媒体影响者的特征对消费者意向及其对购买生酮产品态度的影响》，《国际数据与网络科学期刊》，第6卷，第4期，第1135–1146页，2022年。5. K. L. Lee、P. N. Romzi、J. R. Hanaysha、H. M. Alzoubi 和 M. Alshurideh，《探讨物联网采用的益处与挑战对供应链绩效和组织绩效的影响：马来西亚的一项实证研究》，《不确定供应链管理》，第10卷，第2期，第537–550页，2022年。6. Subramaniam, Arvind. “一种用于图像处理和机器视觉的神经形态方法”，2017年第四届国际图像信息处理会议（ICIIP）。IEEE，2017年 7. Yakopcic, Chris, Raqibul Hasan, and Tarek M. Taha. “用于多层神经网络算法离线训练的忆阻器神经形态电路。” 收录于2015年国际神经网络联合会议（IJCNN），第1–7页。IEEE，2015年。8. T. Sausen 和 A. Liegel，《反洗钱领域的人工智能：变革正在进行》，NICE Actimize，美国新泽西州霍博肯，技术报告，2020年1月。 [在线]。9.《估算毒品贩运及其他跨国有组织犯罪造成的非法资金流》，联合国毒品和犯罪问题办公室，奥地利维也纳，2011年。10. J. Gao、Z. Zhou、J. Ai、B. Xia 和 S. Coggeshall，《使用机器学习算法预测信用卡交易欺诈》，《智能学习系统与应用期刊》，第11卷，第3期，第33–63页，2019年。

11. Y. Bengio、A. Courville 和 P. Vincent，《表征学习：综述与新视角》，IEEE模式分析与机器智能汇刊，第35卷，第8期，第1798–1828页，2013年8月。Subramaniam, A. 和 Sharma, A., 2019. N2NSkip：使用神经元间跳跃连接学习高度稀疏网络。12. J. Heaton，《预测建模中特征工程的实证分析》，收录于 SoutheastCon 会议论文集，2016年3月，第1–6页。13. A. Coates、A. Ng 和 H. Lee，《无监督特征学习中的单层网络分析》，收录于第14届人工智能与统计国际会议论文集，JMLR Workshop Conference，2011年，第215–223页。14. R. C. Watkins、K. M. Reynolds、R. Demara、M. Georgiopoulos、A. Gonzalez 和 R. Eaglin，《追踪非法所得：探索将数据挖掘技术用作调查洗钱的工具》，《警务实践与研究》，第4卷，第2期，第163–178页，2003年6月。15. T. E. Senator、H. G. Goldberg、J. Wooton、M. A. Cottini、A. U. Khan、C. D. Klinger、W. M. Llamas、M. P. Marrone 和 R. W. Wong，《金融犯罪执法网络人工智能系统（FAIS）：通过大额现金交易报告识别潜在洗钱行为》，《人工智能杂志》，第16卷，第4期，第21页，1995年。16. C. Mead。《神经形态电子系统》。《IEEE会刊》，78 2, 10 17 10 1629 36, 1990 ( ): – . . Henry Markram。《蓝脑计划》。收录于 ACM/IEEE 超级计算会议，SC 2006，2 18 53, 2006 页 纽约，NY，美国，. IEEE, ACM。 . J. Schemmel、J. Fieres 和 K. Meier。《模拟神经网络的晶圆级集成》。收录于《神经网络》，2008年。IJCNN 2008。（IEEE世界计算智能大会）。IEEE国际神经网络联合会议，第431 2 19 438页，2008 2019年6月，– . . Rajitha, K., 九月。基于面部视频光谱反射率的心率测量。收录于2019年IEEE图像处理国际会议（ICIP）（第3362–3366页）。IEEE。20. B Govoreanu、GS Kar、Y-Y Chen、V Paraschiv、S Kubicek、A Fantini、IP Radu、L Goux、S Clima、R Degraeve、N Jossart、O Richard、T Vandeweyer、K Seo、P Hendrickx、G Pourtois、H Bender、L Altimime、DJ Wouters、JA Kittl 和 M Jurczak。10×10 nm2 Hf/HfOx交叉阵列电阻式存储器，具有出色的性能、可靠性和低能耗。《电子器件会议国际技术文摘》，第31 3, 6, 9 21 34页，2011 2020年12月。 . Boppidi, P.K.R., 等。使用忆阻器交叉阵列实现快速独立成分分析，以进行盲图像源分离。《IET电路、器件与系统》，14(4)，第484–489页。22. Gregory S Snider 和 R Stanley Williams。采用现场可编程纳米线互连的纳米/CMOS架构。《纳米技术》，18(3):035204，2007年1月 4 23 . . Tsuyoshi Hasegawa、Takeo Ohno、Kazuya Terabe、Tohru Tsuruoka、Tomonobu Nakayama、James K Gimzewski 和 Masakazu Aono。单个固态原子开关实现的学习能力。《先进材料》，22(16):1831–1834，2010年4月 19 24 4 . 神经形态纳米级忆阻器突触 . T. Serrano-Gotarredona、T. Masquelier、T. Prodromakis、G. Indiveri 和 B. Linares-Barranco。忆阻器脉冲时序依赖可塑性及其变体与脉冲神经形态学习系统。《神经科学前沿》，

7 2 2013 4, 7, 8 ( ), . .

25 C. Zamarre no-Ramos、L.A. Camu nas-Mesa、J.A. P erez-Carrasco、T. Masquelier、T. SerranoGotarredona 和 B. Linares-Barranco。关于脉冲时序依赖可塑性、忆阻器器件及构建自学习视觉皮层。《神经科学前沿》，5，2011年。4, 7

战略威慑研究，空军大学，2020年。26. J. F. Weaver，《美国人工智能监管》，载于《人工智能法律研究手册》，爱德华·埃尔加出版社，2018年，第155–212页。27. J. Whittlestone 和 S. Clarke，《人工智能对社会与伦理的挑战》，《牛津人工智能治理手册》，第1卷，第1期，第1–20页，2022年，doi: 10.1093/ox-fordhb/9780197579329.013.3。28. M. Shamout、R. Ben-Abdallah、M. Alshurideh、A. Kurdi 和 H. B.，《S.（2022）。供应链与物流行业采用自主机器人的概念模型》，《不确定供应链管理》，第10卷，第2期，第577–592页。29. S. M. Butt，《2型糖尿病的管理与治疗》，《国际计算机、信息与制造期刊》，第2卷，第1期，2022年。30. H. M. Alzoubi、G. Ahmed 和 M. Alshurideh，《产品质量维度对提升订单赢家和客户满意度影响的实证调查》，《国际生产力与质量管理期刊》，第36卷，第2期，第169–186页，2022年。31. H. M. Alzoubi、H. Elrehail、J. R. Hanaysha、A. Al-Gasaymeh 和 R. Al-Adaileh，《新冠肺炎危机期间供应链整合与敏捷实践在缩短交货周期方面的作用》，《国际服务科学、管理、工程与技术期刊》，第13卷，第1期，第1–11页，2022年。32. M. T. Alshurideh 等，《用于解决供应链管理问题的模糊逻辑辅助人力资源管理》，《运筹学年鉴》，第1–19页，2022年。33. B. Kurdi、H. Alzoubi、I. Akour 和 M. Alshurideh，《区块链与智能库存系统对供应链绩效的影响：来自零售业的实证证据》，《不确定供应链管理》，第10卷，第4期，第1111–1116页，2022年。34. Rajitha, K., 2019年。《真实环境下通过面部视频估算心脏脉搏》。35. B. Linares-Barranco 和 T. Serrano-Gotarredona。《在自适应异步脉冲神经形态纳米技术系统中利用忆阻》。载于《纳米技术》，2009年。IEEE纳米技术第9届会议论文集，第601–604页。IEEE，2009年。8 36. S.R. Deiss、R.J. Douglas 和 A.M. Whatley。《用于神经形态系统的脉冲编码通信基础设施》。载于 W. Maass 和 C.M. Bishop 编，《脉冲神经网络》，第6章，第157–78页。MIT出版社，1998年。8 37. E. Chicca、A.M. Whatley、P. Lichtsteiner、V. Dante、T. Delbruck、P. Del Giudice、R.J. Douglas 和 G. Indiveri。《多芯片脉冲神经形态基础设施及其在方向选择性模型中的应用》。《IEEE电路与系统汇刊I》，5(54)：981–993，2007年。38. G. Eason、B. Noble 和 I. N. Sneddon，《关于涉及贝塞尔函数乘积的某些 Lipschitz-Hankel 型积分》，《伦敦皇家学会哲学汇刊》，A247卷，第529–551页，1955年4月。