arXiv:2304.01328v1 [cs.CV] 3 Apr 2023

廉价伪造检测大挑战

Duc-Tien Dang-Nguyen[1, 3], Sohail Ahmed Khan[1], Cise Midoglu[2],Michael Riegler[2],Pål Halvorsen[2],Minh-Son Dao[4]

[1]卑尔根大学（UiB）、Simula数字工程[2]大都会中心（SimulaMet）、[3]克里斯蒂安尼亚大学学院、[4]日本信息通信研究机构（NICT）{ductien.dangnguyen, sohail.khan}@uib.no、{cise,michael, paalh}@simula.no、dao@nict.go.jp

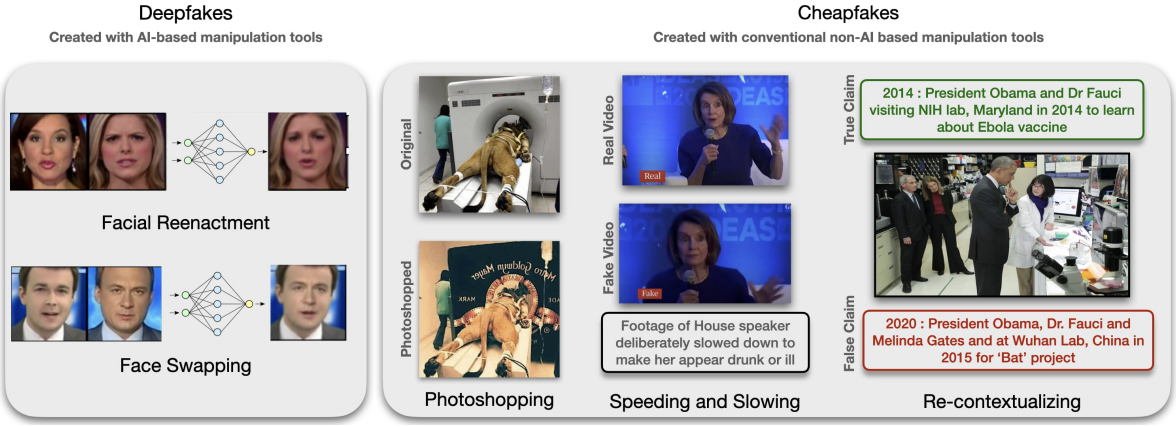


图1：深度伪造（左）是指使用复杂的基于人工智能的媒体操纵工具和技术制作的虚假媒体。廉价伪造（右）包括使用或不使用当代非人工智能编辑工具制作的虚假媒体，这些工具易于获取。图像处理工具可用于篡改图像。视频可以加速或减速，以改变其意图或歪曲视频中人物的形象。重新语境化包括将虚假或不相关的说法与真实图像联系起来，从而歪曲事件或人物。本挑战聚焦于检测经过重新语境化的廉价伪造。图像来源： [1, 2, 3, 4, 5]

1 动机与背景

廉价伪造 是近来新造的一个通用术语，涵盖了不使用深度学习方法创建的非人工智能（“廉价”）多媒体内容操纵。尽管过去几年里，人们对深度伪造的制作、检测和滥用给予了大量关注，但事实上，廉价伪造比深度伪造更为普遍 [6, 7]。

廉价伪造可以借助当代编辑工具制作，这些工具不基于人工智能且易于获取，例如 Adobe Photoshop 或 Premiere Pro；甚至也可以不使用任何软件。图像操纵、加快或放慢视频速度，以及有意篡改多媒体素材的语境，例如在新闻图像说明中通过附上误导性说法来分享媒体，都是目前采用的一些手段（见图 1）。后一种做法称为语境篡改，或对媒体进行脱离语境（OOC）滥用。脱离语境的媒体比伪造媒体更难检测，因为图像和/或视频并未遭到篡改。 [8] 报告了目前在互联网上出现的不同类型廉价伪造的概述。

根据廉价伪造的类型，可以使用不同的检测工具。研究人员已探索用于检测图像操纵（如 Photoshop 修图和图像拼接）的方法 [9, 10, 11, 12]。然而，重新语境化或 OOC 滥用相对小众，尚未得到充分研究；这类做法会将虚假或无关的说法与真实图像关联起来，以歪曲事件或人物。Aneja 等人 [13] 最近提出了这一任务，提供了一个名为 COSMOS 的真实世界新闻帖文数据集，并提出了一种检测廉价伪造的方法，该方法使用 COSMOS 数据集进行了基准测试。

本次挑战旨在基于 COSMOS 数据集的一个版本，开发能够检测语境外图像的模型；更具体地说，是检测新闻报道中图像说明与真实照片相矛盾的滥用情况。我们曾在 2021年ACM多媒体系统会议（MMSys’21）和 2022年ACM多媒体会议（ACMMM 2022）上组织过类似的廉价伪造检测挑战 [14, 15]。

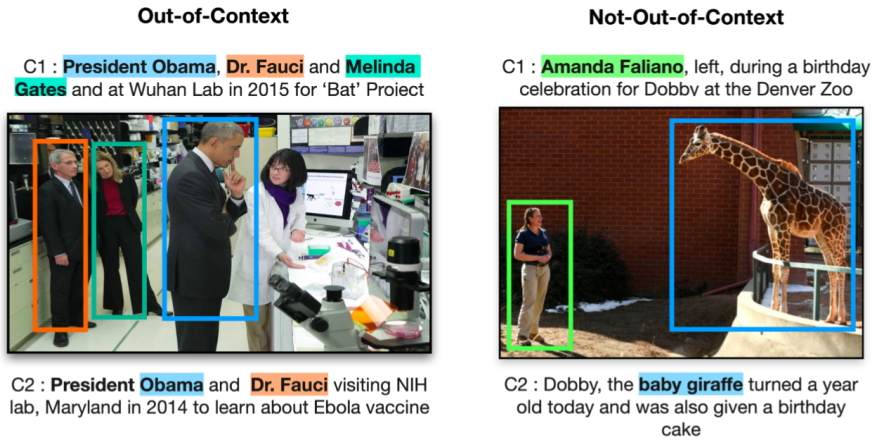


图 2：COSMOS 数据集中的每张图像都配有两条图像说明，它们曾与该图像一同在互联网上传播。左：由于语境被篡改，两条图像说明中的一条具有误导性，表明该图像遭到脱离语境（OOC）滥用。右：两条图像说明均不具有误导性，因此属于未脱离语境（NOOC）。图像来源： [13]

2 主办机构和协调联系人

本次挑战由挪威的卑尔根大学（UiB）、挪威的 Simula 数字工程大都会中心（SimulaMet）以及日本信息通信研究机构（NICT）合作举办。组织委员会：

- Duc Tien Dang Nguyen，UiB，ductien.dangnguyen@uib.no
- Sohail Ahmed Khan，SFI-MediaFutures - UiB，sohail.khan@uib.no
- Cise Midoglu，SimulaMet，cise@simula.no
- Michael Riegler，SimulaMet，michael@simula.no
- Pål Halvorsen，SimulaMet，paalh@simula.no
- Minh-Son Dao，NICT，dao@nict.go.jp

3 挑战描述

图像可作为新闻图像说明所描述事件的证据。将图像用作不真实和/或无关事件的证据，称为图像的脱离语境（OOC）使用（见图 2）。本挑战旨在开发并评测可用于检测新闻报道中图像遭到脱离语境（OOC）误用的模型。

任务 1：检测相互冲突的图像-字幕三元组

如果与某张图像相关联的两条图像说明都成立，那么它们应描述同一事件。如果它们指的是图像中的相同对象，但语义不同，也就是说，将同一主体与不同事件联系起来，这就表明图像被脱离语境（OOC）使用。然而，如果两条图像说明对应的是同一事件，无论它们描述的是哪些对象，这都定义为图像的未脱离语境（NOOC）使用。

在此任务中，参与者需要提出检测相互冲突的图像-字幕三元组的方法，这类三元组表明存在语境错置。具体而言，给定 <Image,Caption1,Caption2> 三元组作为输入，所提出的模型应预测相应的类别标签 1（OOC）或 0（NOOC）。目标不是判断单条图像说明是真实还是虚假，而是检测是否存在语境错置。这类方法被认为特别有助于事实核查人员，因为标出相互冲突的图像-字幕三元组可以帮助他们缩小搜索范围。

任务二：虚假字幕检测

在实际场景中，某张图像可能没有多条图像说明可供参考，因此挑战归结为判断与图像关联的单条图像说明是否真实。

在这项任务中，参与者需要提出方法，判断给定的图像-字幕对是真实的（真实）还是虚假生成的（虚假）。更具体地说，给定一个 <Image,Caption>对作为输入，所提出的模型应预测相应的类别标签：0（真实）或 1（虚假）。

我们承认，即使是人工审核员，在不了解图像来源的情况下，这也是一项具有挑战性的任务。事实上，Luo 等人 [16] 通过一项针对人工评估员的研究证实了这一挑战；研究要求评估员不得使用搜索引擎，而人工准确率平均仅约为 65%。

4 数据集

本次挑战赛将使用经过扩充的 COSMOS 数据集 [13] 。该数据集的一部分经抽样后被指定为公开数据集。公开数据集由训练集、验证集和公开测试集组成，向参与者开放，用于训练和测试其算法。COSMOS 数据集的其余部分将加入新样本并进行修改，以创建隐藏测试集，与 [14, 15] 类似。隐藏测试集不会公开提供，而是由挑战赛组织者用于评估提交结果。挑战赛数据集的详细信息汇总于表 1。

表 1：挑战赛数据集统计信息。

数据集划分	图像数量	图像说明数量	上下文标注
训练集	161752	360749	No
验证集	41006	90036	No
公开测试集	1000	2000	Yes
隐藏测试集	1000	2000	Yes

5 评估标准

参赛模型将根据两个综合得分进行评估和排名，两个得分分别由5项和3项指标组成。

- 有效性：准确率、精确率、召回率、F1分数和马修斯相关系数（MCC）。参赛者需计算其模型的这5项指标，并在提交材料的结果部分列出数值。
- 效率：延迟、参数数量和模型大小。参赛者需计算其模型的这3项指标，并在提交材料的结果部分列出数值。

参赛者在公开测试集上评估自己的模型后，需向组织委员会提供代码和训练后的模型权重，以便在隐藏测试集上评估其模型。参赛者可按第7节所述的三种方式之一提交解决方案，但须遵守第6节所列的截止日期。

6 重要日期

- 数据集发布（公开训练集）：2023年1月16日，星期一
- 数据集发布（公开测试集）：2023年2月20日，星期一
- 模型提交截止日期：2023年3月29日，星期三
- 论文提交截止日期：2023年3月17日，星期一
- 模型评估结果公告：2023年4月3日，星期一
- 论文录用公告：2023年4月24日，星期一

7 提交指南

Docker容器

我们建议挑战赛参与者将解决方案作为 Docker容器提交，以确保不会因软件不兼容问题或其他类似原因而导致错误。在这种情况下，我们建议他们遵循 https://github.com/detecting-cheapfakes/detecting-cheapfakes-code 下给出的说明。

标准Python可执行程序

如果参与者在以 Docker容器 的形式提交解决方案时遇到困难，或者觉得以标准Python项目的形式提交更方便，也可以按照以下说明进行提交。这也

会对我们有所帮助：如果参与者希望提交 Python 项目，最好使用 PyTorch 作为主要库，但这并非强制要求。

• 我们希望提交的代码能够通过一条命令运行，例如：python solution.py <包含隐藏测试集文件的文件夹路径 private_test.json>

• 参与者应确保 private_test.json 文件和 public_test.json 文件采用相同的格式。

• 为应对任何软件不兼容问题，我们要求参与者在提交解决方案的同时提供 requirements.txt 文件，其中列出所用软件包的名称和具体版本号。使用 **Conda** （conda list）和 **Pip** （pip list）都可以轻松完成此操作。我们建议参与者使用 **Conda** ，并在开始编写挑战解决方案的代码之前创建一个全新的环境。

Jupyter Notebook

参与者 也可以按照以下说明使用 Jupyter Notebook 提交解决方案。 .

• 参赛者应合理组织代码，使我们只需修改主文件中的一行即可更改输入路径，即将提交的笔记本中的 INPUT_FOLDER 参数从公开测试集文件的路径更改为私有测试集文件的路径后，即可执行该笔记本，如下所示：

INPUT_FOLDER = <包含隐藏测试集文件 private_test.json 的文件夹路径>

• 参赛者应知悉，private_test.json 文件与 public_test.json 文件的格式相同。

8 附加信息

经验

本挑战赛的前两届分别在以下会议期间举办：(1) 在2021年ACM多媒体系统会议期间，以“MMSys’21 廉价伪造检测大挑战”为名举办 [17, 14], ；(2) 在2022年ACM多媒体会议期间，以“ACM多媒体廉价伪造检测大挑战”为名举办 [15]。

参赛者支持

我们已在 https://detecting-cheapfakes.github.io/ 上建立挑战赛网站，提供有关数据集、任务及其他资源的信息。GitHub 组织 https://github.com/detecting-cheapfakes/ 收录了所有相关代码库。我们还建立了 Google 群组 https://groups.google.com/g/grandchallenge-cheapfakes，为有意参赛者提供支持。感兴趣的参赛者可在这些平台上查阅此前提出的问题并参与互动讨论。

宣传

我们计划通过以下渠道宣传本次挑战赛：(i) 专业网络，例如 MediaFutures、挪威人工智能学会 (NAIS) 和挪威人工智能社区 (NORA)；(ii) 邮件列表、Slack 工作区和公告栏，包括 VisionList 和 Image World；(iii) 本次挑战赛组织者担任组委会成员的会议，例如 ACM ICMR、ACM MMSys、Global Fact、IEEE CVPR、MediaEval、MMM、NAIS 和 NTCIR。

后续

挑战赛网站及其他参赛者资源均已建立，并承诺至少在未来 3 年内持续维护。我们想再次重申，未来几年我们将继续举办这项挑战赛，因为我们认为它既具体、相关且适时，又足够通用，能够随着时间推移根据不同需求不断发展。例如，随着合成数据在研究和训练中的重要性日益提高，未来可能设置的一项任务是生成虚假图像说明。

参考文献

[1] A. Rössler、D. Cozzolino、L. Verdoliva、C. Riess、J. Thies 和 M. Nießner。《FaceForensics++：学习检测经过操纵的面部图像》，2019 年。[2] Boredpanda。《30 张人们信以为真的虚假热门照片》，2019 年。[3] J. Waterson。《Facebook 拒绝删除特朗普支持者传播的虚假佩洛西视频》，2019 年。[4] 白宫。《奥巴马总统参观美国国立卫生研究院疫苗研究中心的一间实验室》，2014 年。[5] D. Evon。《这是 2015 年在武汉实验室的奥巴马、福奇和盖茨吗？》，2020 年。[6] J. S. Brennen、F. M. Simon、P. N. Howard 和 R. K. Nielsen。《新冠肺炎错误信息的类型、来源和主张》，2020 年。[7] N. Schick。《不要低估廉价伪造》，2020 年。[8] B. Paris 和 J. Donovan。《深度伪造与廉价伪造：音视频证据的操纵》，2019 年。[9] C. Chen、S. McCloskey 和 J. Yu。《通过相机响应函数分析检测图像拼接》。*2017 IEEE*计算机视觉与模式识别会议（*CVPR*），第 1876–1885 页，2017 年。[10] D. Cozzolino、G. Poggi 和 L. Verdoliva。《Splicebuster：一种新型盲图像拼接检测器》。*2015 IEEE*信息取证与安全国际研讨会（*WIFS*），第 1–6 页，2015 年。[11] M. Huh、A. Liu、A. Owens 和 A. A. Efros。《打击假新闻：通过学习自一致性检测图像拼接》，2018 年。[12] S. Wang、O. Wang、A. Owens、R. Zhang 和 A. A. Efros。《通过编写 Photoshop 脚本检测经过 Photoshop 处理的面孔》。载于 *ICCV*，2019 年。[13] Shivangi Aneja、Chris Bregler 和 Matthias Nießner。《COSMOS：通过自监督学习捕捉脱离语境的错误信息》，2021 年。[14] Shivangi Aneja、Cise Midoglu、Duc-Tien Dang-Nguyen、Michael Alexander Riegler、Paal Halvorsen、Matthias Niessner、Balu Adsumilli 和 Chris Bregler。《MMSys'21廉价伪造检测大挑战》，2021 年。[15] Shivangi Aneja、Cise Midoglu、Duc-Tien Dang-Nguyen、Sohail Ahmed Khan、Michael Riegler、Pål Halvorsen、Chris Bregler 和 Balu Adsumilli。《ACM多媒体廉价伪造检测大挑战》，2022 年。[16] G. Luo、T. Darrell 和 A. Rohrbach。《NewsCLIPpings：自动生成脱离语境的多模态媒体》，2021 年。[17] C. Midoglu 和 S. Aneja。《廉价伪造检测大挑战》，2021 年。