基于深度学习的图像篡改检测技术在数字档案中的应用研究

◆王璐玥

（上海大学图书情报档案系　上海市宝山区　201900）

数字档案馆作为智慧城市建设的重要环节，作为档案永久保存的电子载体，在信息技术手段日益复杂多变的环境下，人工智能的浪潮给数字档案馆带来了机遇的同时也给档案工作者带来了巨大的挑战。不论是扫描类文档，照片还是声像视频类数字档案，图像都是数字档案资源的主要载体，如何保障图像真实性在新时代仍然是档案学者们关注的重点问题。在“互联网+人工智能”发展的大背景下，中国的人工智能已被提升至国家战略高度，深度学习作为人工智能的重要组成部分，其在档案学领域的应用研究相比于其他领域相对较少，因此必须结合新时代的新技术特点加强对数字档案真实性保护研究，确保数字档案的真实性。

一、深度学习在保障数字档案真实性的应用可行性

（1）技术可行。进入 21 世纪，我国人工智能技术进入蓬勃发展时期。更多的人工智能与智能系统研究课题获得了各种国家基金计划支持，并与中国国民经济和科技发展的重大需求相结合，力求为国家科技发展做出更大贡献。近两年来，中国的人工智能已发展成为国家战略。国家最高领导人习近平、李克强多次发表重要讲话，表示对发展中国人工智能和机器人学给予高屋建瓴的指示与支持。结合深度学习的人工智能在计算机视觉领域更是得到了长足的发展，因此，利用深度学习结合计算机视觉相关领域技术为图像类数字档案篡改检测提供了技术保障。

（2）实践可行。我国人工智能技术攻关和产业应用近年来发展势头迅猛，已经涉及到国民经济 39 个行业的大类，目前已被广泛应用于语音识别、计算机视觉、机器人、语言处理等领域，并且我国目前技术创新能力不断增强[1]。此外，我国数字档案馆建设经过近十年的发展，大部分档案馆信息化基础设施完备，专业技术人力资源充足，国家政策上也给予了一定的支持，相继出台了国家级别的标准 GB/T18894–2002《电子文件归档与管理规范》，行业标准 DA/T15–1995《磁性载体档案管理与保护规范》等我国数字档案资源安全标准[2]，这些都为结合深度学习技术保障图像类数字档案真实性提供了实践可行性。

（3）经济可行。2012 年以来，我国的信息化发展进入新阶段，云计算，物联网，大数据，人工智能等各项技术蓬勃发展，国家特别是中央政府各部门按照国家电子政务相关规划的要求，逐步实施了“一站，两网，四库，十二金”等重点工程[3]。这些国家重点项目的启动为实施深度学习技术在档案领域的应用提供了经济上的支持。

二、深度学习在保障数字档案真实性的应用必要性

（1）图像篡改威胁数字档案真实性。原始凭证性是档案的基本属性[4]，档案一旦失去了真实性的保障，不论从何种角度来说，都会给社会造成难以估量的损失和影响。在互联网即将进入 5G 时代的背景下，信息传播速度快、规模大、影响范围广泛，一旦发生数字档案信息安全问题，相关信息会借助互联网快速扩散出去，势必会引发“蝴蝶效应”[5]，这严重阻碍了数字档案资源建设工作，也使得档案的公信力得不到保障。

（2）对图像类档案管理理论提供支持。图像类档案真实性保障作为数字档案管理的重要环节，数字档案信息安全保障是档案事业现代化、信息化进程中不可避免的问题，也是数字档案馆建设工作中必不可少的环节。数字档案信息安全保障的初衷是为了保证数字档案信息的安全、系统、可用，这是一项复杂而系统的任务，因为数字档案馆建设的每一环节都或多或少存在安全隐患，所以要求数字档案馆建设过程中的每项工作都要小心谨慎、认真仔细。结合深度学习的图像类数字档案篡改检测技术可以作为一项辅助手段，对图像类档案管理提供支持。

（3）多种技术保障图像类档案真实性检测效果。对图像类档案真实性检测的技术补充。目前多数的图像像档案真实性检测都采取依靠数字水印，数字签名等技术的主动取证技术，该类技术有长达几十年相对成熟的发展基础，但其缺点在于必须事前知晓原始水印或者原始签名信息才能进行图像类档案的篡改检测。利用神经网络对图像类档案进行真实性检测属于被动取证技术，不需要提前知晓原始图像信息，仅根据图像本身的纹理特征，结构特点等判断图像类数字档案是否被篡改。

（4）对数字档案馆知识产权的保护。“互联网+”使得数字档案资源的利用范围更加广泛，面临的风险更大。如，数字档案资源知识产权风险，在网络环境下未经允许擅自复制数字档案信息资源的问题较为普遍，这些问题容易引发知识产权纠纷；信息篡改风险，数字档案资源在“互联网+”时代容易被越权非法篡改，影响数字档案的真实性[6]。“互联网+”时代给数字档案资源安全带来了更大挑战，引入图像类档案检测技术，可以助力数字档案安全工作开展，以“互联网+”理念做好数字档案资源建设工作，并为这项工作构建一个安全、平衡的秩序。

三、深度学习在保障数字档案真实性的应用

深度学习（Deep Learning）是一类算法集合，是机器学习的一个分支。它尝试为数据的高层次摘要进行建模。深度学习是一种机器学习架构，所有的个体单元以权重的方式连接在一起，且这些权重是通过网络来训练的，那么它就可以称之为神经网络算法。神经网络算法的思想来源于模仿人类大脑思考的方式。人类大脑是通过神经系统得到输入信号再作出相应反映的，而接受外部刺激的方式是用神经元接受神经末梢转换的电信号。深度学习算法通过采用层的方式组织神经元，层与层之间可以互相连接来模拟大脑的思考。

（1）应用背景

对于图像类数字档案，为了达到隐藏某些信息的目的，主要有复制和粘贴篡改和拼接篡改两种篡改手段[7]。在复制粘贴篡改方法中，篡改者往往会把本图像中的某一个区域复制粘贴到本图像的另一个区域。在拼接篡改中，篡改者为了达到某种目的，把图像 A 的一部分拼接到图像 B 上。为了使篡改不被发现，篡改者往往还会对所要复制的区域进行变换，如旋转和缩放，然后将所变换的区域粘贴到其他区域，并对这个粘贴后的区域进行加噪和模糊等后续处理。这些处理即使在图像表面拼接肉眼无法识别出与真实图像的区别，但在图像的一些低级特征和统计学特征上，不可避免的留下篡改痕迹。这类篡改与真实图像之间区别性的特征可以作为神经网络的输入，利用深度神经网络学习的能力，判别图像的真伪。以图像的边缘为例，篡改图像的粘贴区域，会引起边缘灰度的剧烈变化。如下图所示，在数据库 MICC–F2000 中：图 a，b 为真实为篡改图像，图 c,d 为复制粘贴篡改图像（红色边框内为篡改粘贴区域），图像 a,c 经过边缘提取得到图片 b,d，从图 b,d 的对比中可以看出，篡改粘贴区域面源明显灰度变化明显。

(a)　　(b)

(c) (d)

（2）应用流程

在完整的篡改检测流程中，主要分为以下三个部分：特征提取，训练网络和验证与测试。针对图像篡改检测任务的深度神经网络可以看作完成二分类问题的分类器，输出结果为 1 或 0，即篡改或真实。在特征提取模块，主要是利用数据库训练集里真实图像与篡改图像之间的区别性特征，作为深度神经网络的输入来不断训练网络形成新的参数。在训练网络模块，利用验证集验证训练好的网络是否具备一定的泛化性，如验证集效果不如人意，则需要重新调整神经的参数重复训练过程。在验证集的结果达到一定精度要求后，利用测试集测试网络性能。最终测试集输出的结果即为深度神经网络判别结果。

（3）应用难点分析

国内外已经有开源的针对不同格式图像的篡改检测数据库供研究者训练和测试神经网络，但是在数字档案领域，目前还没有统一的数据库来验证一个算法的好坏。对于研究者来说，第一步是需要建立适用于图像类数字档案领域的篡改数据库。由于训练深度神经网络需要大量的正负样本，因而在采集源数据，制作篡改样本这两项工作上要耗费大量的时间和精力。

从硬件上来说，训练深度神经网络需要高性能的 GPU 进行并行计算，我国目前的数档案馆建设工作虽然推进多年，在档案馆硬件资源上也投入了大批资金，但是在计算机硬件这方面，重视程度还不及档案馆其他硬件设施投入。因此，要引入深度学习技术仍需领导决策层面的重视与支持。

从人力资源配备来说，我国目前的数字档案馆从业人员以档案专业的工作者为主，计算机，统计学等理工科专业人员为辅。将深度学习的方法灵活运用到数字档案领域需要技术人员具有先进的计算机专业知识和深厚的数学功底，这对于数字档案馆从业人员的也是一项新的挑战。

四、总结与展望

现如今科学技术发展日新月异，图像篡改手段日益多样化，篡改图像在表面上看来虽然掩人耳目，但是在检测与篡改的博弈中，只要充分利用现有的技术手段，定能为图像类数字档案真实性保驾护航。作为档案工作者，在夯实自身档案专业素质的同时，也要紧跟技术潮流，及时了解最新技术，探索档案管理新技术。

参考文献：

[1]国务院．国务院关于印发新一代人工智能发展规划的通知．国发〔2017〕35 号：http://www.gov.cn/zhengce/content/2017-07/20/content_5211996.htm，2017-07-20.

[2]张勇．数字档案信息资源安全保障体系研究[D]．苏州：苏州大学，2007:17-19.

[3]中国电子政务网.我国电子政务工程的顶层设“两网一站四库十二金”.http://www.e-gov.org.cn/egov/web/article_detail.php?id=151138，2014-07-29.

[4]徐峰．试论档案在依法治县中的作为[J]．档案记忆，2015(7):23-23.

[5]聂云霞，张加欣，甘敏．“互联网+”背景下数字档案资源安全研究[J]．浙江档案，2016(6):22-25.

[6]许鹏．“互联网+”时代数字档案信息资源建设探讨[J]．城建档案，2017(10):30-31.

[7]李子健．图像盲篡改检测算法研究[D]．北京：北京交通大学，2017：11-13.

注释：

①复制粘贴篡改图像数据库 MICC-F2000，源网址：http://lci.micc.unifi.it/labd/2015/01/copy-move-forgery-detection-and-localization/

作者简介：王璐玥，女，1994 年 12 月，民族：汉，江苏盐城人，硕士学位，上海大学，研究方向：数字档案馆。