



QUANDO A EVIDÊNCIA É UMA MIRAGEM

O título dessa reportagem, "Quando a evidência é uma miragem", na interpretação de um programa de IA

Sistema em desenvolvimento recorre à inteligência artificial para detectar adulterações em imagens de artigos científicos

Fabrcio Marques

Um consrcio de pesquisadores do Brasil, dos Estados Unidos e da Itlia est desenvolvimento um conjunto de ferramentas computacionais para detectar de forma automtica a existncia de adulteraes ou duplicaes de imagens de artigos cientficos, um tipo de m conduta muito frequente cuja identificao ainda hoje depende, em boa medida, de um olhar humano experiente. Em um trabalho publicado no final de outubro na revista *Scientific Reports*, do grupo Nature, a equipe apresentou uma primeira avaliao de desempenho do Sila (Anlise de Imagens Cientficas). Trata-se de um sistema de apoio a revisores e editores de revistas acadmicas capaz de processar arquivos em formato PDF de artigos, extrair suas imagens de forma automtica – rastreando eventualmente cpias em alta resoluo disponibilizadas pelos autores ou as publicaes – e em seguida utilizar algoritmos treinados para identificar eventuais adulteraes.

Como resultado, o software informa se encontrou evidncias de “ps-processamento” de imagens, tais como indcios de que regies foram clonadas ou movidas de lugar ou de que uma foto tem similaridades com outras j publicadas anteriormente, apontando tambm, em formato grfico, qual a sua provenincia. O Sila segue um modelo cooperativo entre inteligncia artificial e humana. Ainda que aponte para contedos suspeitos automaticamente, sua apreciao a apenas indicativa: cabe a um especialista verificar os indcios e confirmar se eles configuram mesmo adulterao. A anlise foi feita com base em material colhido em 988 artigos cientficos que sofreram retratao, nos quais foram documentadas manipulao ou reutilizao de figuras. “A ferramenta se baseia em princpios avanados de processamento de imagem, em tcnicas de percia forense e em solues de viso computacional e inteligncia artificial para fornecer anlises que ajudem os especialistas humanos a decidir se os eventos descobertos so le-

gítimos ou não”, afirma um dos autores, Anderson Rocha, docente e diretor do Instituto de Computação da Universidade Estadual de Campinas (Unicamp), onde coordena o Laboratório de Inteligência Artificial (Recod.ai).

Na avaliação divulgada no trabalho, os módulos que compõem o sistema exibiram desempenho desigual. Em tarefas relacionadas à classificação do conteúdo dos *papers*, por exemplo, a resposta dos algoritmos desenvolvidos revelou-se ainda insuficiente. Em outras, como detectar regiões com sinais de deslocamento de imagem e captar indícios de duplicação, alcançou resultados mais sensíveis do que os obtidos pela observação humana, gerando inclusive falsos positivos. “Continuamos a trabalhar no aperfeiçoamento do Sila, pois há bastante terreno para melhorá-lo”, explica Rocha.

Os algoritmos e as bases de dados utilizados foram disponibilizados em repositórios de acesso aberto e a ideia é que outros pesquisadores ajudem a testar e refinar o desempenho da ferramenta. O consórcio tem financiamento da Agência de Projetos de Pesquisa Avançada em Defesa (Darpa) e do Escritório de Integridade Científica (ORI), ambos dos Estados Unidos. No Brasil, é apoiado pela

FAPESP por meio de um projeto temático liderado por Rocha. “Idealizamos o sistema pensando nas necessidades do ORI, que nos fez a encomenda”, diz ele. A ideia é que o ORI e a Darpa avaliem o conjunto de ferramentas em instituições públicas e estabeleçam parcerias com editoras de revistas acadêmicas.

O autor principal do estudo, Daniel Moreira, fez doutorado na Unicamp sob orientação de Rocha e é pesquisador das universidades de Notre Dame e Loyola Chicago, ambas nos Estados Unidos. Todas as etapas do trabalho foram supervisionadas por Edward Delp, professor de ciência da computação, da Universidade Purdue. Pesquisadores de instituições italianas, como a Universidade Nápoles Federico II e o Instituto Politécnico de Milão, também fazem parte do consórcio. A iniciativa integra um projeto mais abrangente que busca criar algoritmos de inteligência artificial para detecção de “realidade sintética”, ou seja, para identificar falsificações em fotos, áudios e vídeos de todo tipo (ver Pesquisa FAPESP nº 321).

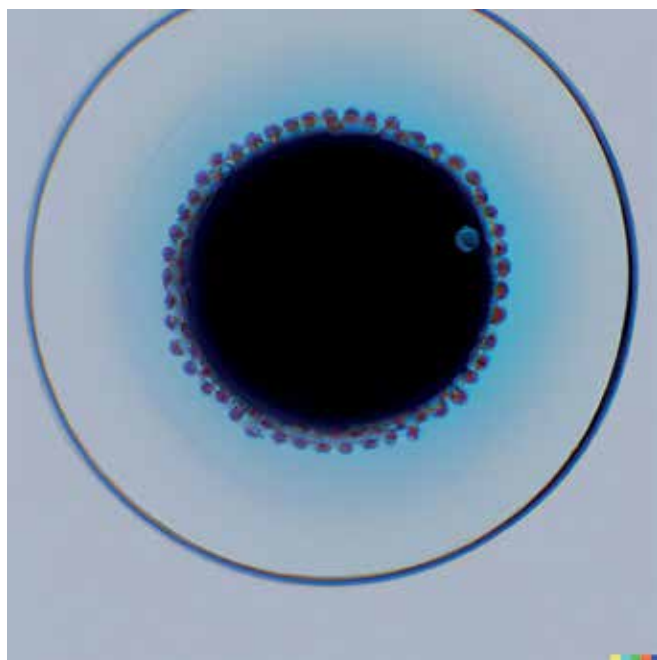
Uma vantagem do Sila é que ele abrange um ciclo completo de tarefas para analisar imagens, da extração à análise, e pretende se tornar uma ferramenta aberta e disponível para todo tipo de usuário. Nos últimos tempos, surgiram no mercado soluções tecnológicas dessa natureza, mas elas buscam resolver tarefas específicas e são comercializadas por empresas.

PRÓXIMO PASSO É CRIAR ALGORITMOS QUE RECONHEÇAM ARTIGOS FEITOS POR “FÁBRICAS DE PAPERS”

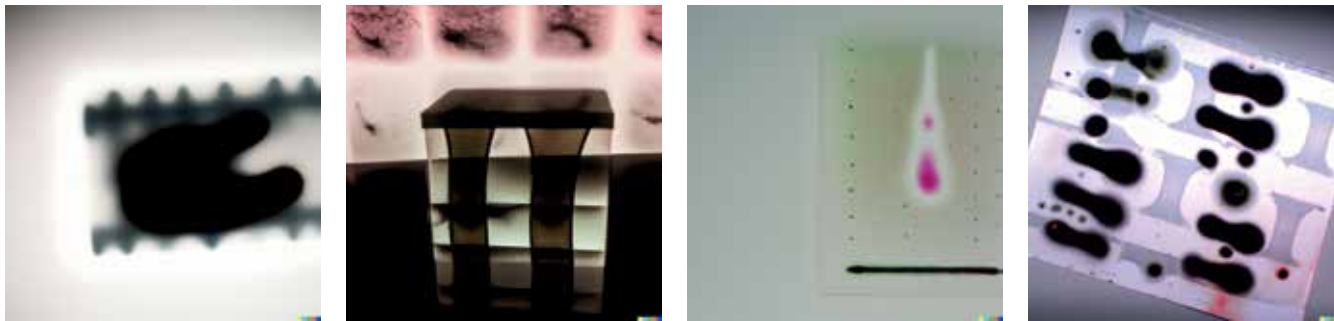
Uma delas é o Image Twin, um software oferecido por uma startup de Viena, na Áustria, que detecta reuso e duplicação de figuras em artigos. Dez revistas editadas pela Associação Americana para Pesquisa do Câncer (AACR) e dois periódicos vinculados à Sociedade Americana para Pesquisa Clínica vêm testando, até aqui com bons resultados, um programa da empresa israelense Rehovot, criado para identificar fotos duplicadas ou com partes giradas, invertidas e esticadas. “Estamos bastante satisfeitos com os resultados até agora”, disse à revista *Nature* Daniel Evanko, diretor de operações da AACR.

Desde o advento da fotografia, imagens são utilizadas como evidências de resultados científicos. Em alguns casos, elas exprimem o próprio achado: caso da fotografia feita pela química britânica Rosalind Franklin, que revelou a estrutura do DNA em 1952 e embasou o artigo assinado por James Watson e Francis Crick. Com a transição da fotografia analógica para a digital, ferramentas de software de edição como o Photoshop permitiram que pesquisadores retocassem arquivos com bastante facilidade, destacando áreas de interesse tais como “bandas” em resultados de testes de *western blot*, método usado na biologia molecular para identificar proteínas. “A maioria dessas edições é legítima. Não há problema em calibrar a intensidade ou o contraste para tornar o resultado mais fácil de visualizar”, afirma Rocha. “Mas há edições que comprometem a integridade e, às vezes, isso tem claramente a intenção de enganar os leitores. Nesses casos, é preciso corrigir o registro e até mesmo pedir a sua retratação.”

Esses problemas são mais frequentes do que se imagina. Uma análise manual realizada pela microbiologista Elisabeth Bik em 2016 vasculhou mais de 20 mil



"Imagem manipulada de célula, em estilo minimalista", em versão do software DALL-E



Opções em estilo surrealista criadas por programa de inteligência artificial a partir do comando "Imagem forjada de teste western blot"

estudos da área biomédica e encontrou algum tipo de adulteração em 4% deles (ver Pesquisa FAPESP nº 310). É fácil enxergar mudanças aberrantes, mas há manipulações sofisticadas que podem tornar a identificação bastante complexa. Um dos principais desafios dos editores atualmente é reconhecer manuscritos produzidos por “fábricas de papers”, serviços ilegais que vendem trabalhos científicos sob demanda, frequentemente com informações falsificadas, inclusive imagens fabricadas. Bik recentemente identificou 400 artigos com figuras tão semelhantes que sugeriam uma origem comum – uma fábrica de papers. Havia indícios, em alguns casos, de que uma imagem padrão foi adaptada para representar diferentes experimentos, moldando-se ao conteúdo de cada documento fraudulento, por meio, por exemplo, de alteração de cores ou da mudança de lugar de algum elemento.

Um desafio para detectar manipulações de forma automática é dispor de repositórios com grandes volumes de fotos, que sirvam de referência para rastrear a origem de conteúdo reaproveitado, e também de bancos de dados com manipulações frequentes, que permitam detectar os padrões de fraudes em manuscritos submetidos para publicação. Na análise do texto de trabalhos científicos, esses recursos já estão disponíveis e munciam o funcionamento de soluções antiplágio. O engenheiro da computação João Phillipe Cardenuto, que está sendo orientado por Rocha no doutorado, publicou em agosto do ano passado um artigo na revista *Science and Engineering Ethics* em que reúne exemplos de adulterações de imagens e apresenta uma biblioteca de algoritmos, disponível em acesso aberto, que reproduzem e são capazes de identificar duplicações, re-

toques e remoções. As imagens disponíveis na biblioteca não são reais, mas, sim, reproduções criadas para simular as características mais comuns das manipulações. Isso porque ele não poderia usar as figuras fraudulentas obtidas em estudos retratados por razões legais e restrições de direitos autorais.

Agora, Cardenuto trabalha para identificar características de imagens adulteradas em documentos gerados por fábricas de papers – e a treinar algoritmos para rastreá-los na literatura acadêmica. Já existem ferramentas automatizadas para detectar artigos desse tipo, como o Papermill Alarm, desenvolvido pela empresa de serviços de dados acadêmicos Clear Skies, no Reino Unido. Ele usa técnicas de aprendizagem profunda (*deep learning*) a fim de analisar títulos e resumos de manuscritos e comparar sua linguagem com a de trabalhos sabidamente produzidos por esses serviços ilegais. O software não indica que o manuscrito é fabricado, mas sinaliza aqueles que merecem uma investigação mais aprofundada antes da publicação. “Como o conteúdo parece fazer sentido, há sempre dúvidas se foram mesmo fabricados. Se pudermos reunir evidências relacionadas a imagens, ficará mais fácil ter certeza”, diz Cardenuto.

O futuro tende a ser ainda mais complexo. Uma fronteira tecnológica envolve a geração de imagens sintéticas, produzidas por programas de inteligência artificial. “Se já pode ser difícil confirmar que uma foto contém elementos de outra, será preciso criar novos parâmetros para verificar fraudes em registros que forem totalmente fabricados”, diz Rocha. Ainda não há provas de que isso esteja acontecendo

em artigos acadêmicos, mas a tecnologia é uma realidade palpável – programas como o DALL-E 2, da OpenAI, é capaz de inventar imagens sobre qualquer tema com base em descrições apresentadas na forma de textos – como se fosse um retrato falado. “A inteligência artificial pode ajudar a detectar dados duplicados em pesquisas, mas também pode ser usada para gerar informações falsas”, escreveu Elisabeth Bik em um texto publicado em outubro no jornal *The New York Times*. “Hoje em dia é fácil produzir fotos ou vídeos fabricados de eventos que nunca aconteceram e imagens geradas por inteligência artificial podem já ter começado a envenenar a literatura científica. À medida que a tecnologia se desenvolve, será significativamente mais difícil distinguir o falso do real.”

Não é impossível rastrear artigos cujos textos foram gerados por programas como os da linhagem do ChatGPT. “Há uma premissa na ciência forense, segundo a qual qualquer intervenção deixa algum vestígio”, afirma Anderson Rocha. “Um algoritmo de inteligência artificial seria capaz de identificar um paper gerado por outro algoritmo de inteligência artificial, mas provavelmente precisará ser aperfeiçoado de forma contínua para acompanhar a sofisticação crescente nas formas de adulteração. Será um jogo de gato e rato para ver quem está mais avançado: nós ou os fraudadores.” ■

Projetos

1. Déjà vu: Coerência temporal, espacial e de caracterização de dados heterogêneos para análise e interpretação de integridade (nº 17/12646-3); Modalidade Projeto Temático; Pesquisador responsável Anderson Rocha; Investimento R\$ 1.912.168,25.
2. Filtragem e análise de proveniência (nº 20/02211-2); Modalidade Bolsa de Doutorado; Bolsista João Phillipe Cardenuto; Pesquisador responsável Anderson Rocha; Investimento R\$ 130.935,02.