



## Integridade Científica - Análise de Fraudes em Imagens de Artigos Científicos

João Phillipe Cardenuto\*, Anderson Rocha

### Resumo

Recentemente, com a quantidade de artigos científicos publicados sendo cada vez maior, o número de casos de má conduta na comunidade científica está crescendo, consumindo investimentos, desgastando profissionais em pesquisas irreprodutíveis e lesando uma das poucas fontes fidedignas de informação na era da pós-verdade, a ciência. Motivado a protegê-la, esse trabalho visa auxiliar a comunidade científica, verificando a integridade de artigos, pela automatização de processos forenses em suas imagens, através de métodos de inteligência artificial e visão computacional. Em seu fluxo de automatização, propomos um processo que se inicia com um artigo digital em formato PDF (padrão na literatura), seguido da identificação e segmentação de imagens passíveis de má conduta e, por fim, na identificação de possíveis atos desonestos em suas imagens. Com auxílio de instituições parceiras - Defense Advanced Research Projects Agency (DARPA) e Universidade de Notre Dame -, realizamos experimentos em artigos da área de Biomedicina e obtivemos resultados impressionantes para detecção de suspeita de fraude.

### Palavras-chave:

Integridade Científica, Forense, Visão Computacional.

### Introdução

Dada a enorme quantidade de artigos científicos publicados diariamente, se tornou inviável a análise manual da integridade científica de cada publicação. Em suporte a esta tarefa, o objetivo dessa pesquisa é automatizar a análise de manipulações por cópia-colagem em imagens de publicações científicas.

A figura 1 mostra o fluxo de execução do modelo proposto, que se inicia com um artigo em formato digital PDF seguido da (1.1) extração de todas suas imagens; (1.2) filtragem de imagens não passíveis de fraude (i.e. logo de artigos, gráficos de barras); (1.3) agrupamento das imagens prováveis de serem cópias uma das outras; (1.4) aplicação de algoritmos de detecção de cópia-colagem da literatura [1], resultando na possível indicação de uma imagem suspeita.

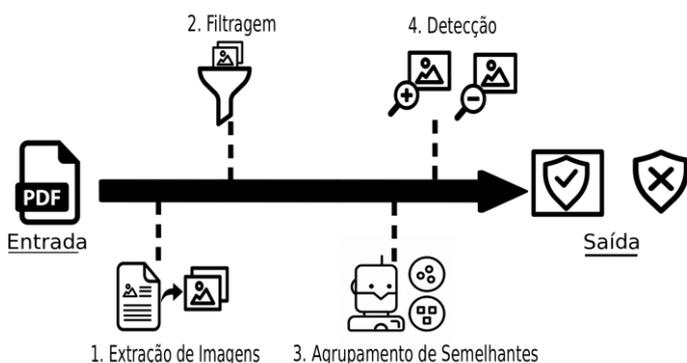


Figura 1. Fluxo de execução do modelo proposto.

### Resultados e Discussão

Visto a carência de modelos propostos para esse objetivo na literatura e ausência de um conjunto de dados para uma avaliação quantitativa, foi feita uma análise qualitativa dos resultados usando como referência artigos da área de biomedicina notificados por manipulações em suas imagens, pelo próprio jornal científico que os publicou.

A Figura 2 mostra um exemplo da execução do modelo proposto, em que a cópia-colagem suspeita é destacada com cores laranja e amarelo por nossa solução. Neste exemplo, o artigo de entrada [2] foi notificado por seu publicador devido a duplicação de imagens, na mesma região destacada por nosso algoritmo.

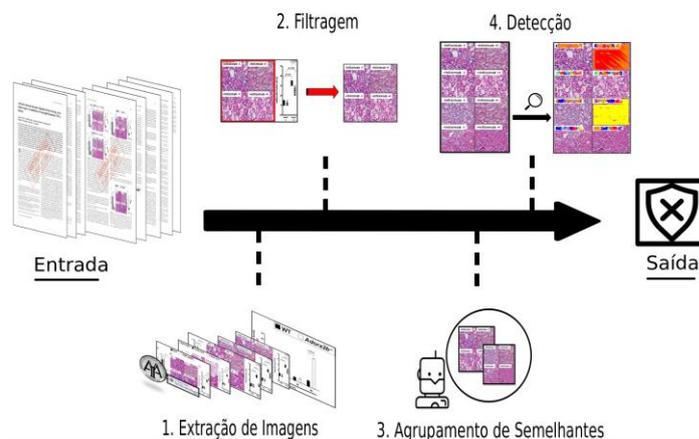


Figura 2. Exemplo de execução do modelo de detecção de cópia-colagem.

Ainda que apresente falsos-positivos, ao destacar em alguns casos palavras de legendas de imagens diferentes (i.e., Figura 2.4), o modelo apresenta de forma qualitativa bons resultados.

### Conclusões

Esta pesquisa mostra a viabilidade para se automatizar o processo de detecção de imagens manipuladas por cópia-colagem em artigos científicos. Para um trabalho futuro, propomos a criação de um conjunto padrão de artigos para se validar quantitativamente modelos de detecção de cópia-colagem em artigos científicos.

### Agradecimentos

FAPESP, DARPA-EUA.

<sup>1</sup> Christlein, V.; Riess, C.; Jordan, J.; Riess, C. e Angelopoulou, E. *IEEE Trans. Inf. Forensics Secur.* **2012**, 7(6), 1841–1854.

<sup>2</sup> Almut G.; Kim J.; Jessica D. B.; Eunyoung T.; Holger K. E. e Eric T. C. *The Journal Of Immunology.* **2012**, 189 (9), 4566-4573.