

TECNOLOGIA

Novo método identifica cerca de 97% da pornografia em telas de celulares e computador

Pesquisadores desenvolvem solução para reconhecer conteúdo pornográfico por meio da combinação de aprendizado de máquina com a análise de imagens estáticas e em movimento

UNICAMP

Um dos problemas mais prementes da universalização do uso do celular e da internet é o acesso indiscriminado a sites e vídeos com conteúdo pornográfico. O acesso a tais conteúdos pode ser voluntário por parte de usuários maiores de idade, mas também pode ser absolutamente desaconselhável quando envolve menores.

O acesso pode ainda ser involuntário, quando se recebe mensagens de correio eletrônico indesejadas, contendo anúncios com conteúdo obsceno, ou ocorrer de forma inadvertida, quando se visita um site invadido por piratas da tecnologia e que lá postaram conteúdo pornográfico.

A indústria de tecnologia da informação e os provedores de conteúdo buscam constantemente alternativas capazes de filtrar em tempo real conteúdo indesejado, bloqueando a sua transmissão ou exibição.

Sistemas do tipo são valiosos para os órgãos policiais, pois aperfeiçoam e agilizam o trabalho de rastreamento e enfrentamento à produção e à divulgação de pornografia infantil, por exemplo. As empresas e escolas também se beneficiam do desenvolvimento de tais tecnologias, pois podem bloquear o conteúdo considerado ofensivo que é acessado por meio de seus computadores.

Foi em busca de uma tecnologia com essas características que pesquisadores do Samsung Research Institute Brazil procuraram em 2012 o Instituto de Computação (IC) da [Universidade Estadual de Campinas \(Unicamp\)](#).

"Eles estavam em busca de uma solução para ser instalada no sistema operacional de celulares, TVs e computadores que permitisse aos consumidores com filhos a possibilidade de bloquear previamente, no momento da compra de um aparelho, por exemplo, o acesso a conteúdo sensível", disse Anderson Rocha, professor do IC-Unicamp e coordenador da pesquisa.

A parceria universidade-empresa levou ao desenvolvimento, em 2015, de um sistema baseado em tecnologia de aprendizado de máquina (ou inteligência artificial), capaz de filtrar mais de 90% do conteúdo pornográfico em um dispositivo. A nova tecnologia foi patenteada em copropriedade entre a Samsung e a Unicamp.

Os pesquisadores prosseguiram buscando novas formas de elevar ainda mais o índice de detecção de conteúdos sensíveis, fossem eles pornográficos ou violentos. A solução que produziu melhores resultados foi combinar informações estáticas de fotografias com informações dinâmicas dos vídeos.

"Temos hoje um filtro bastante eficaz, capaz de identificar mais de 97% do conteúdo pornográfico em geral. Trata-se de um patamar superior às soluções atualmente consideradas como estado da arte, cuja eficácia, muitas vezes, oscila



No novo método, os pesquisadores propõem a combinação do uso de informações estáticas e de movimento com uma metodologia de aprendizado de máquina conhecida como "aprendizagem profunda" (deep learning)

de 87% até 94%", explica o pesquisador, atualmente em período sabático na Nanyang Technological University, em Singapura.

Rocha e outros pesquisadores do IC-Unicamp e da Samsung publicaram recentemente um artigo no periódico *Neurocomputing* no qual detalham os passos do desenvolvimento da nova tecnologia. A pesquisa contou com apoio da FAPESP.

No novo método, os pesquisadores propõem a combinação do uso de informações estáticas e de movimento com uma metodologia de aprendizado de máquina conhecida como "aprendizagem profunda" (deep learning).

"Trata-se de uma metodologia de aprendizado de máquina baseada na aprendizagem de representações não lineares dos dados. Uma imagem, por exemplo, pode ser representada como um vetor de valores denotando diversas transformações lineares nos seus pixels que podem, por exemplo, capturar informações relacionadas à vizinhança dos pixels, formas, arestas, ou de diversas outras formas", explicou Rocha.

Na aprendizagem profunda a ideia é buscar representações melhores em cada nível do aprendizado (normalmente na forma de uma rede com diversas camadas) e criar modelos para aprender essas representações a partir de dados em grande escala. Algumas das representações são inspiradas em avanços em neurociência.

SEXO E VIOLÊNCIA

Os sistemas pioneiros de detecção de pornografia consistem em primeiro tentar detectar cenas de nudez para, em seguida, definir qual o limite de exposição física aceitável, além do qual se configuraria a pornografia, que passaria então a ser filtrada. Essas soluções costumam usar como base de comparação características da pele humana como a cor e

a textura, além de dados da geometria corporal humana.

O resultado muitas vezes deixa a desejar, filtrando menos ou bloqueando mais do que deveriam. O problema é que nem todas as imagens com grande exposição de pele humana têm a ver com sexo, como é o caso de gente se bronzeando ou nadando ou cenas de lutadores de MMA ou de luta greco-romana, por exemplo.

Uma solução mais avançada poderia envolver a filtragem de conteúdo adulto feita com base em uma lista de palavras classificadas conforme descrições do que é permitido e do que é pornográfico.

Tal método consistiria em inserir um estágio intermediário com a descrição da imagem entre a extração inicial de dados do conteúdo a ser filtrado e a sua classificação para permissão ou bloqueio. Entretanto, esse método ainda seria incapaz de distinguir a diferença entre cenas ambíguas. Como diferenciar, por exemplo, um exame médico de pornografia?

No caso de vídeos, os pesquisadores da Unicamp acreditam que se possam reduzir os casos de ambiguidade ao se adicionar outro elemento de classificação: informações de movimento extraídas ao longo do tempo.

A solução que desenvolveram extrai um quadro por segundo de cada vídeo que é acessado em tempo real em celular ou computador. Os quadros com as imagens estáticas são em seguida analisados aplicando-se o método de classificação de descrições do que é permitido e do que é pornográfico.

Ao mesmo tempo, a sequência de quadros analisados fornece os elementos para sequenciar os movimentos dos objetos e pessoas presentes na cena. Dependendo do tipo de movimento, o vídeo é bloqueado.

Segundo Rocha, o método foi

testado em um conjunto de dados contendo aproximadamente 140 horas com mil vídeos pornográficos e mil vídeos não pornográficos que variavam de 6 segundos a 33 minutos.

Os vídeos pornográficos envolviam atores de etnias diversas e também foram considerados desenhos animados. Entre os vídeos não pornográficos havia cenas de banhistas na praia e em clubes, ou combates de lutas. Foi usando essa metodologia que a equipe do Instituto de Computação da Unicamp conseguiu elevar o nível de filtragem de pornografia aos 97%.

"A taxa cai para 90% no conteúdo envolvendo especificamente pornografia infantil e 80% naqueles que têm cenas de violência, situações muito mais difíceis de filtrar. Ambos os temas fazem parte dos novos esforços de pesquisa do grupo do Instituto de Computação da Unicamp. O filtro detecta quando o conteúdo indesejado começa e quando ele termina. O sistema bloqueia a sua exibição assim que a pornografia inicia e volta a liberá-la assim que ela acaba", disse Rocha.

"No contexto forense, o sistema permite que se analise, por exemplo, 30 horas de vídeos no disco rígido do computador de um indivíduo para detectar meia hora de pornografia infantil, configurando assim prova substancial para processá-lo", disse.

A nova tecnologia de análise dos movimentos tem diversas outras aplicações que não só rastrear e bloquear pornografia. Os pesquisadores começam a empregar o método para analisar cenas de violência em manifestações, por exemplo.

"Já é possível rastrear um indivíduo no meio de uma multidão pelo seu modo de andar. É impressionante o que se consegue aprender hoje a partir da análise de dados", disse Rocha.