

Ferramenta PPCensor: detecção de pornografia em tempo real no streaming de vídeo

Jackson Mallmann^{1,2}, Altair O. Santin², Eduardo K. Viegas², Roger R. dos Santos², Jhonatan Geremias²

¹Instituto Federal Catarinense – Jardim Maluche
88.354-300 - Brusque - SC

²Programa de Pós-Graduação em Informática (PPGIA)
Pontifícia Universidade Católica do Paraná (PUCPR)
80.215-901 - Curitiba - PR

jackson.mallmann@ifc.edu.br, {altair.santin, eduardo.viegas, robson.roger, j.geremias}@ppgia.pucpr.br

Abstract. *This paper presents the tool entitled Private Parts Censor (PPCensor) for the detection of pornography-related objects in videos in a network proxy. To achieve such a goal, the tool performs the analysis of the video frames that are currently being downloaded in real time in a transparent manner. For the pornography object detection, an object detector is trained with PPO (Private Parts Object Dataset). When PPCensor identifies a video frame containing private parts (objects), its content is hidden from the current video without the need for user interaction or the requirement of additional processing in the user device. The evaluation results have shown that PPCensor is able to detect private parts in real time during video transmissions.*

Resumo. *Este artigo apresenta a ferramenta Private Parts Censor (PPCensor) para detecção de objetos de natureza pornográfica implementado como um proxy. Para tanto, a ferramenta efetua a análise em tempo real de maneira transparente dos frames de vídeos que estão sendo visualizados por usuários. Para a detecção dos objetos pornográficos, um detector de objetos é previamente treinado com a PPO (Private Parts Object Dataset). Quando PPCensor identifica um frame contendo partes íntimas (objetos), o conteúdo é ocultado do vídeo em exibição sem a necessidade de iteração ou a exigência de processamento no dispositivo do usuário. Os resultados da avaliação mostram que PPCensor é capaz de detectar partes íntimas em tempo real durante transmissões de vídeo.*

1. Introdução

Nos últimos anos, a disseminação de vídeos para maiores de idade na Internet, incluindo vídeos com conteúdo explicitamente pornográfico, aumentou significativamente [Statista 2020]. Por exemplo, em 2018, um dos sites pornográficos mais visitados recebeu mais de 4 milhões de uploads diários de vídeos [Pornhub 2020]. Consequentemente, usuários menores de idade são, de forma prematura e inadequada,

expostos à pornografia. A exposição a esse tipo de conteúdo à menores de idade pode causar constrangimento ou até mesmo traumas psicológicos [PCAA 2020].

Para suprimir esta lacuna, foi desenvolvido a ferramenta PPCensor, implementado através de uma arquitetura para detecção de pornografia em tempo real, onde ocultam-se apenas as áreas inadequadas em um conjunto de *frames*, em vez de bloquear o acesso ao vídeo inteiro. I.e., um servidor proxy de streaming de vídeo. Desta forma, o usuário da ferramenta possui os vídeos com teor pornográfico filtrados durante o processo de visualização da mídia.

A abordagem proposta em PPCensor [Mallmann et al. 2020], a primeira da literatura, realiza a detecção de partes íntimas com taxas de precisão semelhantes em comparação com as abordagens do estado da arte baseadas na classificação de imagens tradicionais. Além disso, para implementação de PPCensor, precisou ser elaborado um *dataset* de objetos íntimos, denominado de PPO (*Private Parts Object Dataset*).

O código fonte, manual e vídeo demonstrativo com a instalação e as funcionalidades de PPCensor estão disponíveis para download¹ e sua publicação aceita e pública [Mallmann et al. 2020] no periódico da *Elsevier Future Generation Computer Systems*. Esta publicação faz parte do projeto CNPQ (processo número 430972/2018-0), intitulado de “Detecção inteligente em tempo real de URLs que disseminam imagem de pornografia infantil em vídeos”. PPCensor [Mallmann et al. 2020] contribui com uma nova arquitetura de detecção de pornografia baseada em CNN (*Convolutional Neural Network*) permitindo que mídias acessadas por dispositivos com recursos limitados sejam monitoradas. A arquitetura é implementada em um servidor proxy junto ao processo de detecção de objeto. Portanto, não demanda modificações e nem processamento extra no dispositivo do usuário, uma vez que PPCensor é executado remotamente e de forma transparente para o usuário final.

O artigo está organizado da seguinte forma. A Seção 2 descreve a ferramenta PPCensor: aspectos técnicos que proporcionam a reprodutibilidade do trabalho. Finalmente, na Seção 3 apresenta-se as conclusões.

2. PPCensor

Nesta Seção apresenta-se a ferramenta PPCensor, tratando dos aspectos relacionados a sua arquitetura e implementação.

2.1. Arquitetura

PPCensor possui uma arquitetura dividida em duas etapas (Proxy e Detector de Objetos de Pornografia) que são aplicadas em um servidor proxy, como ilustrado na Figura 1.

O servidor proxy se interpõe entre os usuários e a Internet [Vicentini 2018]. Possui um cache de navegação sendo capaz de armazenar todas as requisições web provenientes dos usuários em memória. Quando solicitado uma requisição web que já foi carregada, este servidor envia os dados, sem a necessidade de consultar a Internet.

¹ <https://secplab.ppgia.pucpr.br/ppcensor>

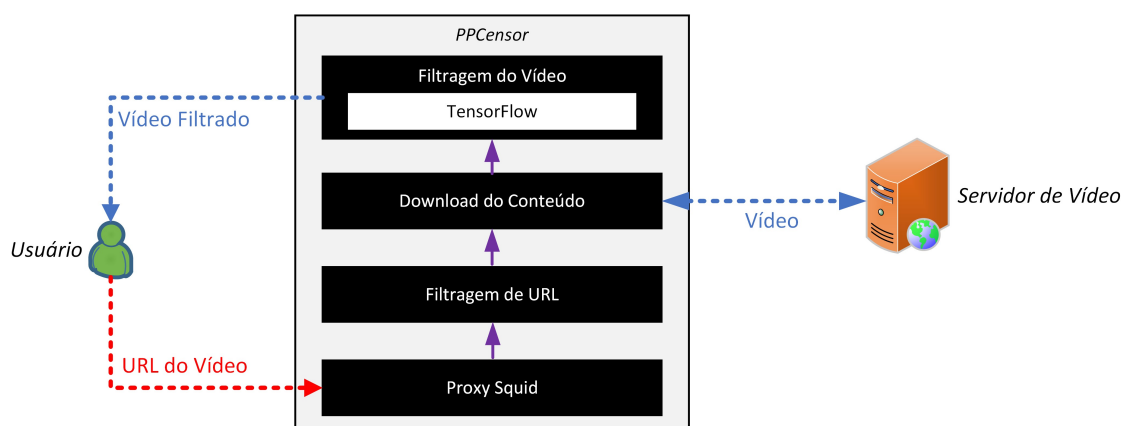


Figura 1. Arquitetura da ferramenta PPCensor

O objetivo é detectar pornografia em tempo real e de fácil uso. Para atingir esse objetivo, PPCensor é executado em um hardware dedicado, porém, de uso geral e a detecção é realizada de forma transparente no dispositivo do usuário final.

O proxy é responsável pela requisição e manipulação de transmissão de vídeo, realizando a análise das requisições do dispositivo do usuário, o download e a filtragem dos vídeos e a devida resposta filtrada relacionada. O processo é executado de filtragem é executado em um servidor proxy. Portanto, o único requisito é que o usuário final altere as configurações do dispositivo para redirecionar consultas para o servidor PPCensor.

A etapa de detecção de objetos de pornografia é responsável por apresentar uma abordagem de detecção amigável. Para tanto, PPCensor é o primeiro trabalho que aborda a tarefa de detecção de pornografia como um problema de detecção de objetos. Isso se baseia na percepção de que partes íntimas possam ser identificadas como objetos. Por exemplo, pode-se desfocar/ofuscar/tarjar as regiões dos *frames* contendo objetos identificados. Como resultado, mesmo que ocorram erros de classificação, a experiência do usuário não é significativamente degradada. Isso ocorre porque o PPCensor trata o conteúdo inadequado como um problema de reconhecimento de objetos, sendo capaz de filtrar os *frames* dos vídeos inadequados em vez de bloquear todo o conteúdo do vídeo.

2.2. Implementação

PPCensor foi implementado em linguagem de programação Python. Inicialmente é descrito a implementação do proxy, e então, do detector de objetos de pornografia.

2.2.1. Proxy

PPCensor foi implementado em um servidor proxy executando o software Squid [Squid 2020] versão 4.5. Esse software está disponível ao público e pode ser utilizado como proxy de rede em servidores baseados em Linux. Quando usuários solicitam requisições de vídeo, o Squid proxy armazena todas as requisições recebidas em um arquivo de log, que é analisado e interpretado por outros processos. Se um vídeo solicitado não se encontra na *Black List*, seu acesso é permitido, para tanto, a URL do vídeo é utilizada para download do vídeo requisitado. Isso ocorre mediante uso da API Pafy [Pafy 2020] versão 0.5.1. Feito o download, executa-se a extração dos *frames* do vídeo, para serem submetidos a segunda etapa (detecção de objetos de pornografia).

Após execução da segunda etapa, executa-se automaticamente o processo de remontagem do vídeo, em que os *frames* são reagrupados a fim de serem enviados para o usuário solicitante fazer a visualização [Viegas 2017]. Essa etapa é executada continuamente e, quando um novo vídeo é baixado, o modelo CNN é aplicado e os objetos de pornografia identificados são rotulados. O módulo também permite adicionar um vídeo a *Black List* do recurso do Squid proxy (Validação Final/Lista de Bloqueio). Portanto, um vídeo classificado como pornografia também pode ser bloqueado em tempo real, se necessário.

2.2.2. Detector de Objetos de Pornografia

Essa etapa fez uso do detector de objetos Faster R-CNN previamente treinado com o COCO *dataset* [Lin et al. 2014]. Este detector se baseia em modelo CNN de dois estágios. Após as características das imagens de entrada serem extraídas por uma CNN, são fornecidas aos dois estágios. O primeiro é a RPN (*Region Proposal Network*): uma CNN que visa propor regiões detectadas, i.e., este módulo determina ao próximo módulo onde deve-se identificar objetos. Para tanto, uma janela deslizante é aplicada na RPN para localização sobre o mapa de características. Por outro lado, no segundo estágio é realizada a detecção, ou seja, para cada localização é verificada a probabilidade da existência do objeto sendo procurado, assim como suas coordenadas [Ren et al. 2017].

Para atingir o objetivo de detectar objeto pornográfico em *frames* foi necessário fazer a obtenção de modelo. Para tanto, utilizou-se a técnica de *transfer learning* [Oquab et al. 2014]. Para tal, utilizou-se do *dataset Private Parts Object Dataset* (PPO) que foi originado no trabalho [Mallmann et al. 2020]. Esses objetos tiveram origem de vídeos do *dataset Pornography-2k* [Moreira 2016]. Durante o treinamento, o modelo que está sendo gerado é validado parcialmente mediante o fornecimento das imagens de validação. Ao final do processo a rede gera o modelo treinado, i.e., um modelo com os pesos ajustados. O modelo treinado é utilizado pelo PPCensor: durante execução do processo, quando um usuário requisita um vídeo ao proxy, este extraí os *frames*, e os envia para a etapa de detecção de objetos de pornografia. Sendo que a segunda etapa é executada no uso da API do TensorFlow Detect Object [TensorFlow 2020] usando o tensorflow-gpu versão 1.14.

3. Conclusões

Este artigo apresenta uma ferramenta social a ser utilizada no consumo de vídeos armazenados na Internet. Uma ferramenta para proxy que analisa/modifica regiões dos *frames* de vídeos pornográficos. As atuais ferramentas publicamente disponíveis para a detecção de conteúdo pornográfico no streaming de vídeo não facilitam a detecção transparente e amigável. Neste artigo, apresenta-se o PPCensor, que se baseia em duas ideias em relação à detecção de conteúdo pornográfico. Primeiro, o PPCensor é o primeiro trabalho que trata o conteúdo pornográfico como um problema de detecção de objetos. Consequentemente, fornece detecção amigável. A avaliação manual de mais de 50 mil *frames* contendo partes íntimas revelou que a técnica proposta pode detectar conteúdo inapropriado em tempo real. Além disso, gera resultados semelhantes às propostas simples para tarefas de classificação de imagens. Ademais, a ferramenta PPCensor é implementada como um servidor proxy; portanto, não incorre em processamento adicional no dispositivo do usuário já que existe execução transparente.

Esta ferramenta poderia ser modelada e utilizada na detecção de PI (Pornografia Infantil), fornecendo as URLs suspeitas para instituições como a NCMEC (*National Center for Missing Exploited Children*). Atualmente, a ferramenta, o código fonte, e o tutorial de instalação encontram-se disponíveis em <https://secplab.ppgia.pucpr.br/ppcensor>.

Agradecimentos

Os autores agradecem ao CNPq pelo apoio financeiro parcial ao projeto, processo: 430972/2018-0, a NVIDIA Corporation pela doação de uma GPU Titan-XP usada nos experimentos, e a URCOP (Unidade de Repressão aos Crimes de Ódio e à Pornografia Infantil). Jackson Mallmann agradece ao IFC (Instituto Federal Catarinense) e ao Capes pela bolsa concedida através do edital n° 231/2017.

Referências

- Lin, T.; Maire, M.; Belongie, S.; Bourdev, L.; Girshick, R.; Hays, J.; Perona, P.; Ramanan, D.; Zitnick, C.L.; Dollár, P. (2014). “Microsoft COCO: Common objects in context”. In ECCV, 2014.
- Mallmann, J.; Santin, A.O.; Viegas, E.K.; Santos, R.R. DOS.; Geremias, J. (2020). “PPCensor: Architecture for real-time pornography detection in video streaming”. *Future Generation Computer Systems*, v. 112, p. 945-955, 2020.
- Moreira, D.; Avila, S.; Perez, M.; Moraes, D.; Testoni, V.; Valle, E.; Goldenstein, S.; Rocha, A. (2016). “Pornography classification: The hidden clues in video space-time”, *Forensic Science Int.* Vol. 268, p. 46-61.
- Oquab, M.; Bottou, L.; Laptev, I.; Sivic, J. (2014). “Learning and transferring mid-level image representations using convolutional neural networks” in *Proc. of the IEEE conf. on computer vision and pattern recognition*, pp. 1717–1724.
- Pafy. (2020), “Library to download YouTube content and retrieve metadata”, <https://pythonhosted.org/pafy/>.
- PCAA. (2020), “Prevent Child Abuse America”, <https://preventchildabuse.org/>.
- Pornhub. (2020), “2018 Year in Review”, <https://www.pornhub.com/insights/>.
- Ren, S.; He, K.; Girshick, R.; Sun, J. (2017) "Faster r-cnn: towards real-time object detection with region proposal networks", *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 39, no. 6, pp. 1137-1149.
- Statista. (2020), “How Much of the Internet Consists of Porn?”, <https://www.statista.com/chart/16959/share-of-the-internet-that-is-porn/>.
- Squid. (2020), “Squid Development Projects”. <http://devel.squid-cache.org/>.
- TensorFlow. (2020), “Tensorflow Object Detection API”, <https://www.tensorflow.org/>.
- Vicentini, C.; Santin, A.; Viegas, E.; Abreu, V. (2018) A machine learning auditing model for detection of multi-tenancy issues within tenant domain. *IEEE International Symposium on Cluster, Cloud and Grid Computing, CCGRID*, p. 543–552.
- Viegas, E.; Santin, A. O.; Franca, A.; et al (2017). Towards an energy-efficient anomaly-based intrusion detection engine for embedded systems. *IEEE Transactions on Computers*, v. 66, n. 1, 2017.