

# VideoRM - Video Maturity Rating

## Detecção de conteúdo impróprio em vídeo

### EQUIPE

#### Coordenador:

Sérgio Colcher (PUC-Rio)

#### Pesquisadores:

Álan Guedes (TeleMídiaPUC-Rio)

Antonio Busson (TeleMídiaPUC-Rio)

Gabriel Noronha (TeleMídiaPUC-Rio)

Pedro Almeida (TeleMídiaPUC-Rio)

### SITE

[www.telemidia.puc-rio.br/projects/videomr.html](http://www.telemidia.puc-rio.br/projects/videomr.html)

### CONTATO

[info@telemidia.puc-rio.br](mailto:info@telemidia.puc-rio.br)



## DESCRIÇÃO

### Introdução e objetivos

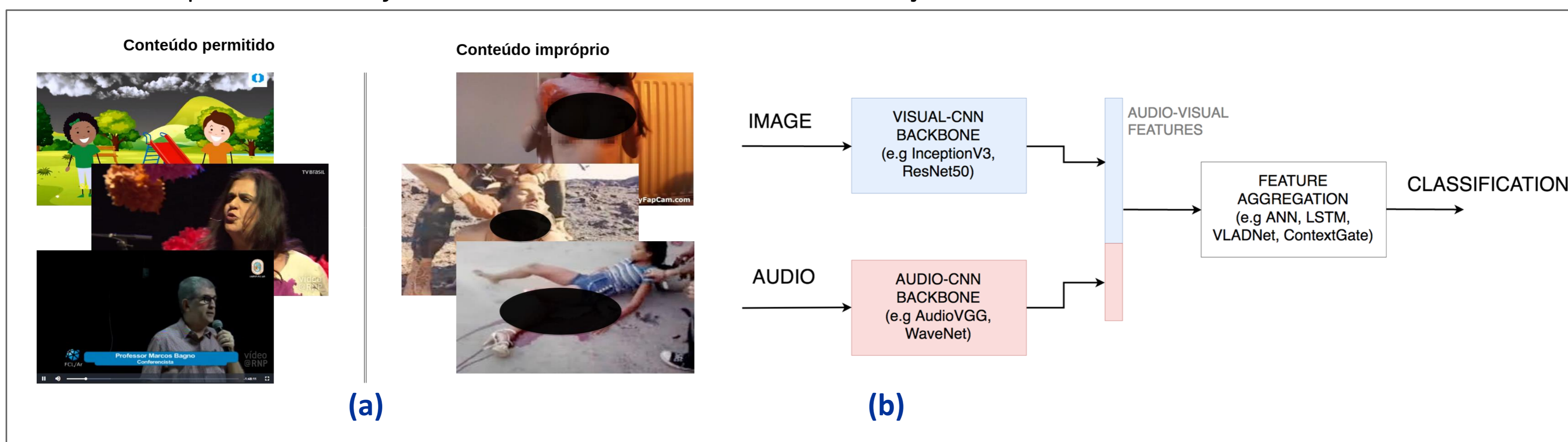
A popularização de equipamentos de captura de vídeo e serviços para seu armazenamento e transmissão possibilitou a produção de um massivo volume de dados de vídeo. O Youtube, por exemplo, registrou em 2014 uploads de 72 horas de vídeo por minuto. Enquanto que em 2018, esse número subiu para 400 horas de vídeo por minuto. Controlar o tipo de conteúdo que é carregado para os serviços de armazenamento é um desafio novo imposto por esse cenário. Por exemplo, serviços como o video@RNP, videoaula@RNP, RUTE e ITVRP constituem redes de compartilhamento vídeo que se focam em conteúdo educacional e possuem restrições à conteúdos impróprios. O carregamento de conteúdo impróprio para essas plataformas por um usuário mal intencionado pode acarretar inclusive problemas jurídicos. Para exemplificar, citamos o caso recente relativo aos assassinatos na Nova Zelândia, em que serviços como Facebook e Youtube estão sendo processados por terem hospedados vídeos dos assassinatos.

A classificação de conteúdo impróprio requer uma análise automática desse volume de vídeos de forma eficiente e prática. Métodos baseados em *Deep Learning* (DL) se tornaram o estado-da-arte em vários segmentos relacionados a análise automática de mídia. Este projeto tem como foco avaliar e desenvolver tais métodos de DL para detecção de conteúdo impróprio em cenas de vídeo.

### Metodologia

Consideramos como conteúdo impróprio cenas de vídeo que estão na faixa 18+ do Sistema de Classificação Indicativa Brasileiro, ou seja, vídeos que contenham conteúdos como violência de forte impacto (e.g. tortura, mutilação e *gore*) e sexo ou nudez sensualizada (e.g. sexo explícito e pornografia). A Figura 1 ilustra as duas categorias de vídeo reconhecidas pela ferramenta de classificação. Em (a) são ilustradas três cenas com conteúdo permitido e (b) são ilustradas três cenas com conteúdo impróprio

O projeto propõe utilizar uma arquitetura bimodal. Primeiro as CNNs, chamadas de *backbones*, serão usadas para extrair as *features* audio-visuais dos frames e áudio do vídeo de datasets de conteúdo próprio e impróprio. Para extração das *features* visuais serão avaliadas o uso das CNNs Inception e ResNet. Já para extração de *features* de áudio, será avaliado o uso das CNN adaptadas AudioVGG e WaveNet. Após a extração, métodos sofisticados para agregação de *features* como LSTM e VLADNET são avaliados para minerar as *features* audio-visuais e realizar a classificação.



(a) Comparação entre conteúdo próprio e impróprio. (b) Arquitetura bimodal com agregação de features visuais e de áudio.

