

# RECONHECIMENTO FACIAL INFANTIL EM IMAGENS: PARA FUTURAS DETECÇÕES DE CONTEÚDOS ILÍCITOS

Gabrielle Mitie Suzuke Tengan<sup>1</sup>, Leila Cristina Bergamasco<sup>1</sup>

<sup>1,2</sup> Departamento de Ciência da Computação, Centro Universitário da FEI

[unifgtengan@fei.edu.br](mailto:unifgtengan@fei.edu.br) [leila.cristina@fei.edu.br](mailto:leila.cristina@fei.edu.br)

**Resumo:** O objetivo do presente projeto é otimizar o tempo de análise de evidências que envolvem conteúdo ilícito infantil, obtidos pelo Ministério Público Federal (MPF) através de um portal de denúncias. Para isso, foram utilizados três tipos de arquiteturas de redes neurais convolucionais – LeNet-5, AlexNet e VGG-16 – e, também um modelo pré-treinado – YOLOv8. Todos foram testados com uma base de dados pública da plataforma kaggle. Por fim, quanto aos resultados, foi atingido uma acurácia máxima de 89,9% com o modelo YOLOv8.

## 1. Introdução

A Internet é um vasto ambiente visto que conta com mais de 64,4% da população [1] mundial ativa, onde grande parte dessa porcentagem são menores de idade. Com isso, é possível perceber que com as diversas opções de entretenimento infantil observa-se a maior vulnerabilidade dos jovens dentro do meio digital, uma vez que, por ser um ambiente amplo, somado à inocência das crianças, sua vulnerabilidade acaba sendo exposta para criminosos da Internet.

Com mais de 23 mil denúncias registradas somente em território brasileiro nos primeiros quatro meses de 2023 [2], é evidente que a pornografia infantil é uma das problemáticas que mais afligem a segurança dos menores em meio digital atualmente. Dito isso, nota-se um aumento considerável das denúncias de pornografia infantil, comparado ao mesmo período do ano de 2022, houve um aumento de 70% [2].

Para a análise dessas denúncias e possível prisão dos criminosos, os especialistas criminais precisam realizar a análise individual desses relatos e categorizá-los de acordo com o crime cometido e se a denúncia possui evidências válidas ou não. O fato de que nem todas as denúncias são válidas é um problema grave, pois acaba gerando a subutilização dos especialistas criminais.

Ao mesmo tempo técnicas de processamento de dados e mais especificamente de imagens, tem evoluído de forma consistente nos últimos anos agregando técnicas de Inteligência Artificial (IA) para o processamento e classificação de imagens em diferentes classes. Essa metodologia pode ser aplicada em diferentes setores da sociedade, como na saúde, na análise de imagens médicas para detectar doenças e na agricultura para identificar tipos de pragas [3]. Na área de segurança pública, essas técnicas podem identificar padrões de imagens ilícitas e bloquear o conteúdo ou avisar autoridades, empresas como Google e Apple já aplicam isso em suas plataformas [4].

Dessa forma, esse projeto de pesquisa busca aplicar e validar técnicas de processamento de imagem e IA para a classificação de faces infantis. Essa primeira

classificação visa facilitar e agilizar o primeiro atendimento de denúncias focadas em pornografia infantil, na medida que poderia filtrar denúncias que efetivamente possuam crianças nas imagens como evidências ou não. A partir dessa filtragem, as imagens com crianças seriam separadas para uma análise futura a ser realizada por um especialista criminal, otimizando o seu tempo para a análise.

Essa pesquisa é parte de um projeto maior que envolve a parceria do Centro Universitário FEI e o Ministério Público Federal de São Paulo (MPF-SP) - divisão de crimes cibernéticos, no qual se planeja implementar algoritmos que auxiliem na classificação de discursos de ódio e pornografia infantil.

## 2. Metodologia

Dado o objetivo do presente projeto, ele foi decomposto em 5 blocos de atividades:

- a) **Pesquisa Bibliográfica:** o aluno pesquisador se aprofundou na literatura sobre conceitos, métodos e documentação necessárias para o desenvolvimento do projeto;
- b) **Formação da base de dados:** na primeira fase, o aluno realizou testes em bases de dados públicas como o [5] da plataforma *kaggle*. A base de dados possui duas classes, adultos e crianças. As imagens possuem resolução de 640x640 e formato RGB.
- c) **Desenvolvimento de técnicas computacionais:** nessa etapa foram implementadas 1) Técnicas de aprendizado de máquina; 2) Implementação de redes neurais convolucionais seguindo as arquiteturas dos modelos LeNet-5, AlexNet e VGG-16, ilustrados nas figuras 4, 5 e 6 respectivamente com base no livro de Ramo [6]; 3) Utilização de um modelo pré treinado.
- d) **Avaliação e Análise:** foram utilizadas as métricas de acurácia, precisão e revocação – que utilizam como base a matriz de confusão - além da validação cruzada.
- e) **Documentação e Divulgação:** a partir dos resultados parciais e totais do presente projeto de pesquisa, relatórios e artigos científicos foram elaborados para divulgação em congressos científicos. Em relação a avaliação serão consideradas métricas de acurácia, precisão e revocação para avaliar o método desenvolvido.

### 3. Resultados

Nas figuras 1 e 2, estão os resultados de acurácia, precisão e revocação de cada modelo, na base de teste e treino, respectivamente. Observando as figuras 1 e 2, é possível notar que, embora o modelo VGG-16 tenha obtido os melhores resultados nas métricas no conjunto de treino, o modelo YOLOv8 obteve os melhores resultados no conjunto de teste. Desse modo, é possível observar que no VGG-16 ocorreu o overfitting – figura 3 – enquanto no YOLOv8 – figura 4 – o modelo conseguiu generalizar a base de dados, atingindo a faixa de 89.9% de acurácia top-1 – métrica de acurácia convencional segundo [7]. Analisando melhor, como o YOLOv8 é um modelo pré treinado com capacidade de classificação, há diversas vantagens em usá-lo, uma delas é a base de dados usada, que é pequena.

| TESTE   | Acurácia                             | Precisão                             | Revocação                            |
|---------|--------------------------------------|--------------------------------------|--------------------------------------|
| AlexNet | $0.594 \pm 0.0314$                   | $0.595 \pm 0.0316$                   | $0.594 \pm 0.0314$                   |
| LeNet   | $0.543 \pm 0.0174$                   | $0.543 \pm 0.0174$                   | $0.543 \pm 0.0174$                   |
| VGG-16  | $0.635 \pm 0.0140$                   | $0.636 \pm 0.0144$                   | $0.635 \pm 0.0140$                   |
| YOLO    | <b><math>0.839 \pm 0.0244</math></b> | <b><math>0.839 \pm 0.0246</math></b> | <b><math>0.839 \pm 0.0245</math></b> |

Figura 1 - Tabela de resultados da base de teste

| TREINO  | Acurácia                        | Precisão                        | Revocação                       |
|---------|---------------------------------|---------------------------------|---------------------------------|
| AlexNet | $0.976 \pm 0.0472$              | $0.977 \pm 0.0460$              | $0.976 \pm 0.0471$              |
| LeNet   | $0.998 \pm 0.0030$              | $0.998 \pm 0.0030$              | $0.998 \pm 0.0030$              |
| VGG-16  | <b><math>1.0 \pm 0.0</math></b> | <b><math>1.0 \pm 0.0</math></b> | <b><math>1.0 \pm 0.0</math></b> |
| YOLO    | $0.895 \pm 0.0175$              | $0.901 \pm 0.0344$              | $0.891 \pm 0.0313$              |

Figura 2 - Tabela de resultados da base de treino

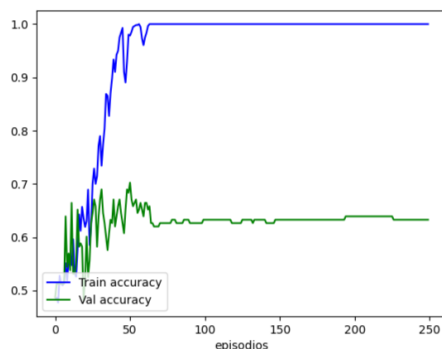


Figura 3 - Acurácia na base treinamento usando VGG-16

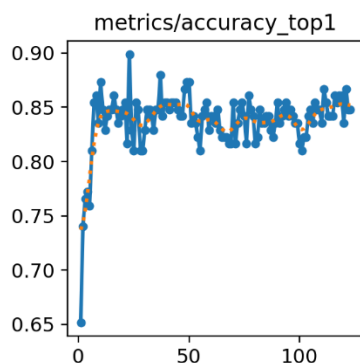


Figura 4 - Acurácia top -1 YOLOv8

### 4. Conclusões

O objetivo de realizar o reconhecimento facial infantil foi atingido com acurácia máxima de precisão de 89.9% utilizando o modelo pré-treinado YOLOv8. Deve-se observar que o modelo pré-treinado se comportou melhor em relação às outras arquiteturas de rede por conta das vantagens que possui.

### 5. Referências

- [1] KYLELORETAN. The Changing World of Digital In 2023-WeAreSocialUK <https://wearesocial.com/uk/blog/2023/01/the-changing-world-of-digital-in-2023/>. [Accessed 02-10-2023].
- [2] REDACÃO, Da. Denúncias de imagens de abuso e exploração sexual infantil online compartilhadas pela SaferNet com as autoridades tem aumento de 70% em 2023-cultura930.com.br. <https://www.cultura930.com.br/denuncias-de-imagens-de-abuso-e-exploracao-sexual-infantil-onlinecompartilhadas-pela-safernet-com-as-autoridades-tem-aumento-de-70-em-2023/>. [Accessed 03-11-2023].
- [3] MARQUES, Bruno Henrique Pereira. Avaliação de algoritmos baseados em deep learning para localizar placas veiculares brasileiras em ambientes complexos. [S.l.: s.n.], 2019.
- [4] CARVALHO, Itamar Almeida de. Classificação de imagens de pornografia e pornografia infantil utilizando recuperação de imagens baseada em conteúdo. [S.l.: s.n.], 2012.
- [5] CHILDREN vs Adults Classification. [S.l.: s.n.]. <https://www.kaggle.com/datasets/die9origephit/children-vs-adults-images>. (Accessed on 03/28/2024).
- [6] RAMO, Klevis. Hands-On Java Deep Learning for Computer Vision. Birmingham, England: Packt Publishing, fev. 2019.
- [7] METRICS - Ultralytics YOLOv8 Docs. [S.l.: s.n.]. <https://docs.ultralytics.com/reference/utils/metrics/#ultralytics.utils.metrics.SegmentMetrics.fitness>. (Accessed on 05/17/2024).

### Agradecimentos

À instituição FEI, seu corpo docente, direção e administração que oportunizaram.

<sup>1</sup> Aluno de IC do Centro Universitário FEI. Projeto com vigência de 04/2023 a 05/2024.