Estimação de idade em imagens digitais a partir de deep learning para apoiar análise pericial

Leandro Amorim Salles[1], **Pedro Luiz de Paula Filho**[1], **Arnaldo Candido Junior**[1]

[1]Universidade Tecnológica Federal do Paraná (UTFPR)
Medianeira – PR – Brazil

leandrosalles@alunos.utfpr.edu.br, pedrol@utfpr.edu.br, arnaldoc@utfpr.edu.br

Resumo. *Com o decorrer da evolução tecnológica de redes sociais e comunidades online, a privacidade e segurança na Internet se tornaram essenciais. O elevado número de informações compartilhadas pela rede sustenta a propagação de conteúdos ilícitos envolvendo pornografia infantil. A visão computacional, fazendo uso de redes neurais como técnica de aprendizado profundo, tem a capacidade de reconhecer características associadas à classificação de conteúdo pornográfico infantil. Aplicadas em imagens digitais e utilizando estruturas neurais pré-treinadas para identificação de rosto infantil em bases de imagens pré-selecionadas, este trabalho teve o objetivo de fornecer subsídios capazes de agregar conhecimento aos métodos de análise pericial. Utilizando a rede neural DenseNet201 para realizar a classificação, o trabalho obteve 96,85% de acurácia em sua performance máxima.*

Abstract. *With the technological evolution of social networks and online communities, privacy and security on the Internet have become essential. The high number of information shared by the network supports the spread of illicit content involving child pornography. Computer vision, based on neural networks as a deep learning technique, can recognize characteristics associated with the classification of child pornographic content. Applied in digital images and using pre-trained neural structures for infant face identification in pre-selected image bases, this work aimed to provide subsidies capable of adding knowledge to the methods of expert analysis. Using the DenseNet201 neural network to perform the classification, the work reached a maximum performance of 96.85% accuracy.*

1. Introdução

A rede mundial de computadores é o principal mecanismo de obtenção de dados dos mais variados tipos. Rapidamente surgiram serviços de compartilhamento de informações entre pessoas ou empresas em diversos níveis, como profissional ou de relacionamento. As Redes Sociais são sites e aplicativos de interação social que se tornaram praticamente indispensáveis à geração nativa do mundo digital e que, recentemente, é alvo de debate em relação à segurança e privacidade de seus usuários que são cada vez mais jovens [Machado 2017].

Devido à falta de regulamentação e precariedade de infraestrutura da rede (segurança, fiscalização, leis internacionais para uso e punição, por exemplo) os crimes cibernéticos associados à pornografia infantil ganharam um grande incentivo

[Hamada and Sanchez 2007, Machado 2017]. O código penal prevê crimes contra a dignidade sexual, possuindo capítulo específico acerca de crimes sexuais contra vulneráveis. O artigo 217-A (Lei n.2.848, de 07 de Dezembro de 1940) define "ter conjunção carnal ou praticar outro ato libidinoso com menor de 14 (catorze) anos: Pena - reclusão, de 8 (oito) a 15 (quinze) anos". Completando, o artigo 218 (Lei n.2.848, de 07 de Dezembro de 1940) define "induzir alguém menor de 14 (catorze) anos a satisfazer a lascívia de outrem: Pena - reclusão, de 2 (dois) a 5 (cinco) anos" [Brasil 1940].

A proteção à criança e ao adolescente situa-se em tratados internacionais, constando no artigo 227 da Constituição Federal. O Estatuto da Criança e do Adolescente[1] (ECA) combate a produção, venda e distribuição de material pornográfico de menores, criminalizando quaisquer tipos de conduta relacionada à pedofilia na Internet. Em [Brasil 1990], como citado no artigo 241-A (Lei n.8.069, de 13 de Julho de 1990), parágrafo único,

> "oferecer, trocar, disponibilizar, distribuir, publicar ou divulgar por qualquer meio, inclusive no meio digital, fotografias, vídeos ou outros registros que contenham cena de sexo explícito ou pornografia envolvendo criança ou adolescente implica pena de 3 (três) a 6 (seis) anos de reclusão e multa".

Por conseguinte, o artigo 241-B (Lei n.8.069, de 13 de Julho de 1990), parágrafo único, define que

> "adquirir, possuir ou armazenar, por qualquer meio, fotografia, vídeo ou outra forma de registro que contenha material pornográfico ou cena de sexo explícito envolvendo criança ou adolescente resulta em pena de 1 (um) a 4 (quatro) anos de reclusão e multa".

O combate ao crime de conteúdo ilícito infantil na Internet é de responsabilidade da Polícia Federal. Para identificar e validar a conduta criminosa, os policiais podem tanto realizar operações para apreensão de computadores ou discos rígidos suspeitos de conter material ilegal quanto instalar programas em servidores responsáveis pelo monitoramento de arquivos suspeitos assim que trafegados pela rede. Essa fiscalização ocorre por meio de *hashes*, assinaturas de um documento que o deixam distinguíveis dos demais existentes na internet [Caiado and Caiado 2018]. Ainda, segundo [Caiado and Caiado 2018], esse procedimento nem sempre é efetivo pois a simples movimentação de arquivos não catalogados previamente em *hashes* exigiria a presença de um perito capaz de analisá-los tecnicamente. Em [Ramos 2018], a perita criminal federal Márcia Mônica Nogueira Mendes constata que a análise técnica visual pode ser extremamente custosa para os profissionais criminais. Além de lidar com um volume grande de arquivos, vídeos e imagens a serem identificadas crianças e adolescentes em cena, é preciso avaliar indícios de produção ou compartilhamento dos arquivos em que a pornografia foi confirmada, representando não somente uma carga psicológica mas também demandando mais tempo para ser concluída.

Ao tratar de reconhecimento de características, principalmente utilizando imagens digitais como objetos de análise, surge uma das principais abordagens computacionais estudadas pela comunidade científica de computação, o denominado *deep learning*. Emergente dentro do campo de Inteligência Artificial, o *deep learning* (aprendizado profundo)

[1] https://presrepublica.jusbrasil.com.br/legislacao/91764/estatuto-da-crianca-e-do-adolescente-lei-8069-90

é, em essência, um modelo matemático de rede neural artificial em que os elementos constituintes (neurônios) são organizados em camadas. Embora frequentemente associado a serviços da computação, o aprendizado profundo também está presente em áreas como a medicina, realizando diagnósticos médicos precisos através da análise de imagens e automobilística por meio de carros autônomos [Ponti and da Costa 2018].

O *deep learning* é considerado o estado da arte relacionado à visão computacional, graças a sua capacidade de permitir o reconhecimento de padrões pelo uso de seus algoritmos conhecidos como redes neurais [Ponti and da Costa 2018]. Este projeto empregou seu uso como técnica para análise de imagens digitais a fim de classificar faixas etárias na tentativa de agregar conhecimento no mesmo segmento utilizado para a realização da análise pericial, como principal alvo a identificação de menores de idade.

2. Redes Neurais Artificiais

Em [Haykin 2008], uma RNA (Rede Neural Artificial) pode ser considerada como um processador paralelamente distribuído formado por unidades de processamento simples, propensos naturalmente a armazenar conhecimento experimental e disponibilizá-lo para uso. [Mitchell 1997] especifica que cada unidade integrante das redes neurais artificiais recebe entradas de valor real (podendo ser saídas de outras unidades) e produz um único valor de saída, também real, podendo se tornar a entrada para posteriores unidades.

As redes neurais convolucionais, conhecidas pela sigla CNN (do inglês *Convolutional Neural Networks*), são modelos aplicados principalmente na área de processamento de imagens digitais com o objetivo de reconhecer padrões [Haykin 2008]. Essa eficácia na classificação de imagens é uma das principais razões do reconhecimento que o mundo deu ao poder do *deep learning* [Hope et al. 2017]. A estrutura de uma CNN possui camadas com neurônios arranjados em três dimensões: largura, altura e profundidade. Essa configuração se encaixa perfeitamente no tratamento de imagens, considerando os píxeis como altura e largura e a informação RGB como profundidade [Raschka 2015, Goodfellow et al. 2016, Patterson and Gibson 2017]. Como demonstrado na Figura 1, os neurônios em uma camada convolucional se conectam à pequenas regiões locais das camadas anteriores. Pode-se resumir que a função de uma camada convolucional é receber e processar informações tridimensionais produzindo resultados de mesma dimensão [Buduma and Lacascio 2017].

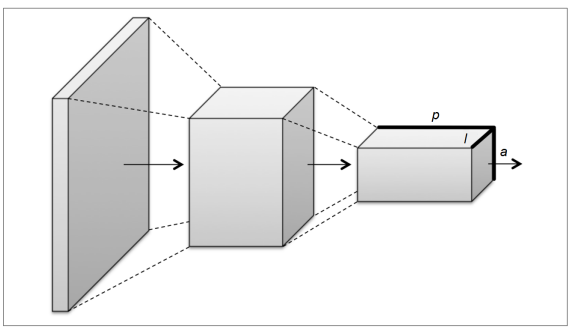

Figura 1. Largura, altura e profundidade de camadas convolucionais.
Fonte: [Buduma and Lacascio 2017].

3. Trabalhos correlatos

[Anand et al. 2017] propôs uma combinação de redes heterogêneas treinadas para extrair recursos de imagens faciais não ideais, ou seja, condições em que ruídos como fundo ou outros objetos estão presentes. Para a estimação de idade, os autores utilizaram o método de regressão aplicado à uma rede *feed forward* para gerar um valor numérico real, que por sua vez foi classificado em um grupo de idades. Os testes ocorreram em três datasets públicos: *WIKI Dataset, AmI-Face Dataset e Adience Benchmark Dataset*. Todos estes compostos por imagens de personalidades públicas em situações diversas de iluminação e cenário. [Anand et al. 2017] obteve uma acurácia final de 58.49% em um máximo de 2.000 épocas, utilizando o Erro Absoluto Médio (MAE) como métrica de validação.

[Rodríguez et al. 2017] utilizou um método conhecido como atenção. Esse mecanismo permite que a rede neural busque por mais detalhes em regiões particulares da imagem de entrada para reduzir a complexidade e, eventualmente, descartar informações irrelevantes. No estudo, os autores utilizaram amostras de faces recortadas e em foco, não tendo sido necessário realizar a identificação durante o processo de treinamento. O modelo CNN foi baseado na estrutura VGG-16, proposto em [Simonyan and Zisserman 2014], com os testes realizados no *Adience Dataset*, que possui uma divisão de 8 faixas etárias com aproximadamente 26.500 amostras. Em 30 épocas os autores conseguiram uma acurácia de 60.78%.

[Rothe et al. 2018] propôs uma abordagem denominada DEX, do inglês *Deep Expectation*, podendo ser traduzida como expectativa profunda. Diferente de alguns estudos, o método utilizado consiste em identificar e recortar a área em que a face aparece em foco de maneira que tanto as imagens de treinamento quanto as de teste possuam a mesma resolução e contenham o mínimo de ruído de fundo possível. A rede convolucional utilizada foi a VGG-16 previamente treinada no conjundo de dados ImageNet. Com o objetivo de aumentar significativamente o conjunto de dados contendo imagens rotuladas com idade, os autores criaram a própria base de dados, o chamado **IMDB-WIKI**, obtendo imagens do IMDb e Wikipédia, totalizando 523.051 amostras de faces. Para fins científicos, os autores tornaram o dataset público. Dividindo em 8 faixas etárias, 0-2, 2-6, 8-13, 15-20, 25-32, 38-43, 48-53 e 60-resto os autores obtiveram uma acurácia média de 64%.

4. Materiais e Métodos

4.1. Ambiente de codificação, linguagem de programação e bibliotecas

O projeto foi desenvolvido utilizando o **Google Colab** como plataforma de codificação. Com uma placa gráfica Tesla K80, o Colab fornece a possibilidade de processar os dados com rapidez, disponibilizando 12GB de GPU. Possui como vantagem a integração nativa da linguagem **Python**, frequentemente usada para a área de aprendizado de máquina e escolhida para a codificação do projeto proposto.

Pela linguagem Python, dois *frameworks* principais foram fundamentais, **TensorFlow** e **Keras**. O primeiro é definido como uma "Biblioteca de software de código aberto para computação numérica de alto desempenho", oficialmente associado à categoria de "*graph compilers*" (computação numérica para gráficos de fluxo). Desenvolvido por pesquisadores e engenheiros de inteligência artificial da Google Brain, oferece recursos fortemente flexíveis para lidar com os processos numéricos dentro do aprendizado da rede

[TensorFlow 2018]. O segundo é uma biblioteca de *deep learning* modulada para Python. Essa biblioteca trabalha com diferentes *backends* para treinamento de redes neurais, incluindo o TensorFlow. Oferece muitas facilidades para criação de modelos. Ainda pelo Keras, é possível importar diretamente várias estruturas neurais pré-treinados, utilizadas no desenvolvimento do projeto.

4.2. Base de dados

Para a tarefa de identificação etária, foi utilizado o *dataset* **UTK Face**[2]. Disponibilizado apenas para fins de pesquisa, esse acervo é composto por mais de 23.000 imagens faciais de indivíduos etnicamente variados. Além da grande quantidade de dados, o UTK possui todas as amostras devidamente rotuladas em cada nome de arquivo no formato "[idade]_[gênero]_[raça]_[data&tempo].jpg".

Para os campos, tem-se:

- "[idade]": número inteiro no intervalo de 1 a 116;
- "[gênero]": 0 para homem e 1 para mulher;
- "[raça]": número inteiro entre 0 e 4, onde denotam, respectivamente, branco, asiático, indiano e outros (hispânico, latino, Oriente Médio);
- "[data&tempo]": formato yyyymmddHHMMSSFFF e mostra a data e tempo em que as imagens foram coletadas.

Originalmente recortadas com o rosto em foco, todas as imagens deste conjunto possuem a mesma resolução de 200x200 píxeis, com algumas amostras demonstradas na Figura 2.

Figura 2. Exemplo de amostras presentes no *UTK Face Dataset*.

4.3. Modelo de rede neural pré-treinada

Para o treinamento de identificação etária, o modelo neural utilizado foi o DenseNet proposto por [Huang et al. 2016], considerado o estado da arte para problemas de classificação até a data em que este trabalho foi desenvolvido. Mais precisamente a versão **DenseNet201**, que alcançou taxa de erro no conjunto de dados ImageNet de 21.46 no top-1 e 5.54 no top-5. A Figura 3 ilustra a arquitetura de camadas da rede.

[2]https://susanqq.github.io/UTKFace/

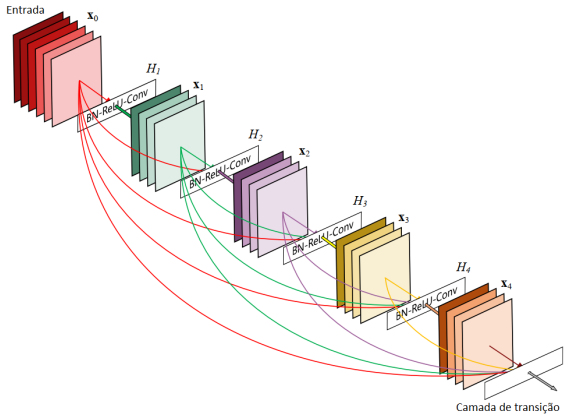

Figura 3. Arquitetura *DenseNet* com 5 camadas.
Fonte: Adaptado [Huang et al. 2016].

4.4. Método proposto

Nosso método consistiu em dividir a base de dados em classes representadas por faixas etárias manualmente organizadas. O *dataset* foi organizado em imagens de treinamento, validação e teste utilizadas pela rede neural para a classificação final. Para a avaliação da proposta foram determinados três experimentos, nos quais os métodos de pré-processamento estão descritos nas seções a seguir.

4.4.1. Experimento 1

O primeiro experimento consistiu em realizar um agrupamento de idades em quantidades de faixas semelhantes à literatura científica do projeto. Para a realização do processo de classificação, as imagens originais foram agrupadas em sete classes: 1-10, 11-15, 16-18, 19-25, 26-40, 41-60 e 61-acima. A Tabela 1 demonstra a quantidade de amostras presentes em cada classe após o agrupamento.

Tabela 1. Distribuição de imagens por cada faixa etária no Experimento 1.

Experimento 1	
Classes	**Nº de amostras**
1-10	3.207
11-15	610
16-18	667
19-25	3.142
26-40	9.363
41-60	4.311
61-r	2.397
Total	**23.697**

Para que a performance de classificação fosse capaz de atingir seu valor máximo, foi realizado um balanceamento das classes menos representadas para as etapas de treinamento e validação do modelo neural. A fim de igualar a representatividade de cada classe,

técnicas de *data augmentation* foram aplicadas às classes 11-15 e 16-18. Realizando um processo denominado *flip-horizontal* (também conhecido como espelhamento), o número de amostras foi dobrado nas duas classes. A Figura 4 ilustra a aplicação do processo em uma amostra.

Figura 4. Exemplo de *flip horizontal*. (a) imagem original, (b) imagem espelhada.

Para evitar realizar demasiadas modificações nas imagens originais e produzir um balanceamento baseado na classe com menos representatividade, um total de 6.748 amostras foram utilizadas, sendo 4.998 (75%) para treinamento, com 714 para cada classe, 1.400 (20%) separadas para validar o processo de treino, sendo 200 amostras como suporte para cada classe e 350 (5%) para teste, sendo 50 amostras em cada classe.

4.4.2. Experimento 2

O segundo experimento teve como principal objetivo verificar a capacidade do modelo neural de diferenciar os indivíduos classificados como vulneráveis perante a lei (de 1 a 13 anos) dos menores de 14 até 17 anos. Assim sendo, foram separadas quatro classes para representar os menores e maiores de idade: 1-13, 14-17, 18-35 e 36-acima. A Tabela 2 representa a distribuição de amostras em cada classe após a divisão do *dataset* original.

Tabela 2. Distribuição de imagens por cada faixa etária no Experimento 2.

Experimento 2	
Classes	**Nº de amostras**
1-13	3.483
14-17	739
18-35	10.874
36-r	8.601
Total	**23.697**

Bem como no Experimento 1, a divisão original não apresentou balanceamento de todas as classes. Possuindo uma representatividade menor, a classe 14-17 também passou por técnica de *data augmentation* no conjunto de treinamento. Conforme ilustrado pela Figura 5, o processo de espelhamento foi novamente utilizado para dobrar o número de amostras. As quantidades nas demais classes foram organizadas baseando-se no número de imagens readequadas com *data augmentation* na classe 14-17, com o objetivo principal de manter o mesmo número de amostras para todas as faixas.

Figura 5. *Data Augmentation* no Exp. 2. (a) original, (b) espelhada.

Para o processo de classificação, foi determinado um total de 4.712 imagens, onde 3.512 (74,5%) foram utilizadas para treinamento, com 878 amostras em cada classe, 800 (17%) para validação do treinamento, sendo 200 amostras como suporte para cada classe e 400 (8,5%) para teste, com 100 amostras em cada classe.

4.4.3. Experimento 3

Por fim, o terceiro experimento buscou simplificar a complexidade de decisão da rede e adequar a classificação para o principal objetivo da análise pericial: diferenciar menores e maiores de idade. Para isso, foram determinadas duas classes: 1-17 e 18-acima. A Tabela 3 apresenta a quantidade de amostras para cada classe no *dataset* original.

Tabela 3. Distribuição de imagens por cada faixa etária no Experimento 3.

Experimento 3	
Classes	**Nº de amostras**
1-17	4.222
18-r	19.475
Total	**23.697**

Com a separação do conjunto de dados em duas faixas etárias, tanto a classe de menores quanto a de maiores de idade possuíam uma quantidade suficiente de amostras, sendo desnecessário realizar processos de *data augmentation*. Ao invés de aumentar o número de amostras da classe 1-17, a quantidade de imagens de 18-r foi reduzida. Dessa forma, o balanceamento binário do *dataset* foi composto por um total de 8.440 imagens, sendo 7.186 (84%) utilizadas para treinamento, com 3.593 amostras para cada classe, 1.000 (11%) para validação, com suporte de 500 amostras para cada classe e 254 (5%) para teste, com 127 amostras em cada classe.

4.4.4. Hiperparâmetros e Métricas de Avaliação

Em relação a todo processo de treinamento dos experimentos propostos, tem-se o uso de parâmetros, funções de perda e otimizadores que auxiliam a rede neural na obtenção da melhor performance de classificação e segmentação. A função de perda responsável por atualizar os pesos da rede em cada retropropagação foi a *categorical crossentropy*

[Goodfellow et al. 2016]. As taxas de aprendizado das estruturas neurais alternaram-se entre 10^{-3} e 10^{-4}. Completando os parâmetros, o tamanho do lote (quantidade por época) de imagens que alimentaram a rede foi de 32.

O modelo neural utilizado para realização dos testes passou por um processo de ajuste fino (também conhecido pelo inglês *fine tuning*), método que tem como objetivo aumentar a performance de generalização da rede durante a fase de treinamento. Como técnica desse ajuste, foi implementada uma camada de *Dropout* com valor de 0.5, proposta em [Hinton et al. 2012], antes da camada de saída completamente conectada.

Para realizar a avaliação dos resultados obtidos pelos experimentos descritos, foram utilizadas como métricas a *acurácia* (ACC) e a *medida-f (f1-score)*. A primeira representa a taxa de acerto global, ou seja, a proporção de predições corretas em relação ao tamanho do conjunto total de dados como mostrado pela Equação 1. Como legenda, v_p refere-se aos verdadeiros positivos, v_n aos verdadeiros negativos, f_p aos falsos positivos e f_n aos falsos negativos [Silva et al. 2012].

$$ACC = \frac{v_p + v_n}{v_p + v_n + f_p + f_n} \tag{1}$$

A segunda configura a média harmônica entre precisão ($Prec$) e sensibilidade (Sen), como apresentado na Equação 2. A precisão determina a porcentagem de amostras que foram classificadas como sendo da classe positiva que de fato a compõe. Por sua vez, a sensibilidade (também conhecida pelo inglês *recall*) indica a proporção de amostras da classe positiva identificada corretamente, mostrando o quão bom o classificador é realizando a tarefa de reconhecer a classe positiva. Essa fusão de métricas torna a *f1-score* uma validação confiável até mesmo para conjuntos de dados desbalanceados.

$$\textit{f1 - score } = 2 * \frac{Prec * Sen}{Prec + Sen} \tag{2}$$

5. Resultados e Discussão

5.1. Resultados do Experimento 1

O Experimento 1 teve como principal objetivo verificar a performance da rede neural em diferenciar uma quantidade diversificada de faixas etárias. A separação visou englobar características faciais semelhantes através da análise do conjunto de dados, definindo as classes 1-10, 11-15, 16-18, 19-25, 26-40, 41-60 e 61-acima. Determinando um total de 50 épocas para o treinamento e definindo a taxa de aprendizado da rede em 10^{-3}, a DenseNet201 obteve um percentual final médio de 62,00% de acurácia. A Tabela 4 demonstra as métricas relacionadas à validação *f1-score* por cada classe do experimento, resultando em um percentual médio de 60,35%.

A Figura 6 apresenta a matriz de confusão para a melhor performance de classificação, no qual pode ser observada a distribuição de acertos e erros ocorridos entre as 7 classes. Como esperado, a diagonal principal da matriz se manteve com os maiores valores, correspondentes aos acertos da classificação, enquanto a confusão da rede entre as faixas etárias ocorreu, em seu maior número, nas classes vizinhas.

Tabela 4. Resultados da métrica *f1-score* do Experimento 1.

Classes	Precisão	Recall	f1-score	Nº de amostras
1-10	71,93%	82,00%	76,64%	50
11-15	60,00%	30,00%	40,00%	50
16-18	50,00%	46,00%	47,92%	50
19-25	52,00%	52,00%	52,00%	50
26-40	55,41%	82,00%	66,13%	50
41-60	66,67%	48,00%	55,81%	50
61-acima	75,81%	94,00%	83,93%	50

	1-10	11-15	16-18	19-25	26-40	41-60	61-r
1-10	41	5	2	0	2	0	0
11-15	14	15	14	5	1	1	0
16-18	2	5	23	15	3	1	1
19-25	0	0	6	26	16	2	0
26-40	0	0	0	3	41	5	1
41-60	0	0	1	1	11	24	13
61-r	0	0	0	0	0	3	47

Figura 6. Matriz de confusão do Experimento 1.

Notou-se que as classes com menor índice de acerto foram as 11-15 (15), 16-18 (23), 19-25 (26) e 41-60 (24). Se analisado em termos de características faciais, é difícil até mesmo para o ser humano determinar com precisão a idade de uma pessoa, principalmente quando os valores etários correspondem à apenas um ou dois anos de diferença. Para a rede neural, foi notoriamente mais difícil perceber as peculiaridades em amostras pertencentes a intervalos de idade próximos em que as características físicas não apresentam mudanças tão visíveis. Por exemplo, indivíduos entre 11-15 anos podem ser mais facilmente classificados como 1-10 ou 16-18 comparado à faixas mais distantes na fase adulta. A Figura 7 demonstra alguns exemplos de amostras do conjunto de teste que foram erroneamente classificadas pela rede.

Figura 7. Exemplos de amostras erroneamente classificadas no Experimento 1.

Como método de análise, adequou-se a matriz à premissa principal de separar menores e maiores de idade, particionando-a em 4 partes principais (Figura 8). As matrizes em verde correspondem aos acertos relacionados às classes agregadas de menores e maiores de idade, enquanto as matrizes em vermelho correspondem aos erros. Para este experimento foram utilizadas 50 imagens por classe como conjunto de teste, isso signi-

fica que 150 amostras validaram a agregação de menores e 200 a agregação de maiores. Somados os valores internos de cada matriz positiva, para a primeira, 121 se encaixaram corretamente na categoria sem levar em consideração os erros internos, representando 80,66% das amostras. Já para a classe de maioridade etária, 193 foram corretamente classificadas gerando percentual de 96,5%. Realizada a média, temos para essa binarização de classes um total de 88,58% de amostras corretamente classificadas.

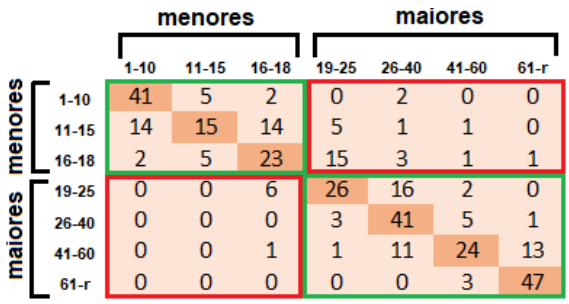

Figura 8. Matriz de confusão do Experimento 1 agregando em menores e maiores de idade.

5.2. Resultados do Experimento 2

A separação das classes para o Experimento 2 teve como inspiração a distinção perante o código penal entre as diferentes idades que compõe crianças e adolescentes. Para as novas faixas, tem-se os considerados indivíduos vulneráveis entre 1 e 13 anos englobados em uma mesma classe, com o restante sendo definido pelas faixas 14-17, 18-35 e 36-acima. De um total de 20 épocas de treinamento e definindo a taxa de aprendizado da rede em 10^{-4}, a DenseNet201 obteve um percentual médio de 73,00% de acurácia. A Tabela 5 demonstra os valores, por classe, de cada métrica relacionada à medida *f1-score*, que obteve um resultado médio de 72,85%, representando um aumento de 12,5 pontos percentuais comparado ao experimento anterior. Por sua vez, a Figura 9 traz a matriz de confusão para a melhor performance de classificação.

Tabela 5. Resultados da métrica *f1-score* do Experimento 2.

Classes	Precisão	Recall	f1-score	Nº de amostras
1-13	73,04%	84,00%	78,14%	100
14-17	64,86%	72,00%	68,25%	100
18-35	68,67%	57,00%	62,30%	100
36-acima	86,81%	79,00%	82,72%	100

Observando a tabela e a matriz correspondente à nova divisão de classes, logo nota-se a grande capacidade de reconhecimento da rede para as classes nos extremos da divisão (1-13 e 36-acima), com os maiores percentuais de precisão e *recall*. Em complemento, também percebeu-se um aumento na proporção média de amostras das classes positivas identificadas corretamente nas faixas intermediárias. De 100 imagens, 72 amostras foram acertadamente atribuídas na classe 14-17, faixa que obteve a menor precisão com 64,86%. Em relação ao Experimento 1, o menor valor de *recall* aumentou em 27 pontos percentuais comparado à classe 11-15, detentora do menor percentual no experimento anterior. Essa análise auxilia a compreender o aumento na métrica *f1-score* geral.

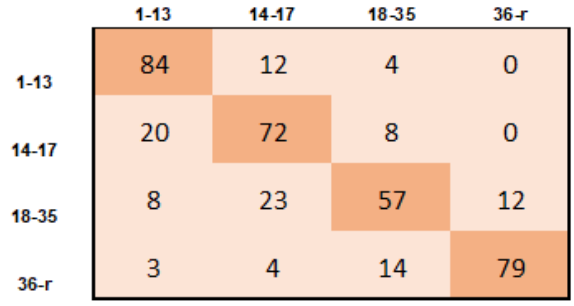

Figura 9. Matriz de confusão do Experimento 2.

Ainda observando a matriz na Figura 9, observa-se mais amostras da classe 14-17 classificadas como 1-13 do que classificadas como 18-35. Assim como no Experimento 1, muitos indivíduos com idades entre essas três faixas podem ser facilmente confundidos baseando-se em características faciais. Entretanto, pelo resultado obtido ficou evidente que, para essa classificação, a confusão ocorreu em maior número para indivíduos mais novos do que em relação aos mais velhos. Por fim, notou-se a qualidade da rede em identificar padrões entre as classes mais distantes, onde nenhuma amostra pertencente à classe 1-13 foi rotulada como acima de 36 anos e somente 3 amostras pertencentes à classe acima de 36 anos foram classificadas como 1-13. Na Figura 10, temos alguns exemplos de enganos cometidos pela rede neural utilizando as imagens separadas no conjunto de teste.

Figura 10. Exemplos de amostras erroneamente classificadas no Experimento 2.

A Figura 11 apresenta a agregação das 4 faixas etárias na premissa de separar em duas categorias representadas por menores e maiores de idade. Sendo 100 imagens utilizadas para teste de cada uma das 4 classes iniciais, em cada matriz positiva destacada em verde são 200 amostras de suporte. Para a agregação de menores de idade, somados os valores internos da nova matriz, 188 imagens foram corretamente classificadas, totalizando 94,00% de precisão e um aumento de 13,34 pontos percentuais comparado à agregação do Experimento 1. Em contrapartida, para a classe de maioridade etária houve decréscimo de 15,5 pontos percentuais, com 162 amostras corretamente identificadas e totalizando 81,00% de precisão. Realizada a média dessa binarização de classes tem-se um total de 87,5%, representando um decréscimo de 1,08 pontos percentuais em relação ao mesmo processo realizado no Experimento 1.

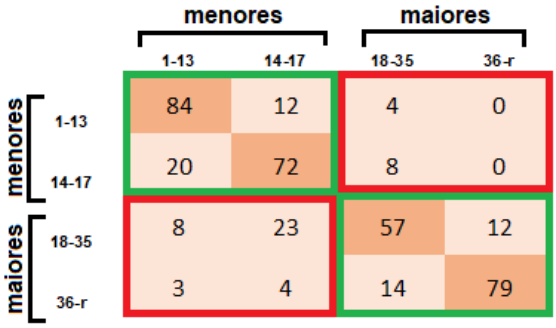

Figura 11. Matriz de confusão do Experimento 2 agregando em menores e maiores de idade.

5.3. Resultados do Experimento 3

Como teste final para identificação de faixas etárias, o Experimento 3 teve como principal objetivo binarizar a classificação da rede neural e propor o treinamento para as classes de menores e maiores de idade, representados por 1-17 e 18-acima. Em um total de 20 épocas e mantendo a taxa de aprendizado da rede em 10^{-4}, a DenseNet201 obteve percentual médio de 96,85% de acurácia, 23,85 pontos percentuais a mais comparado ao experimento anterior.

A Tabela 6 apresenta os resultados das métricas relacionadas à validação *f1-score* por cada categoria. Com um percentual final médio de 96,85%, houve um aumento de 24 pontos percentuais em relação ao Experimento 2. Mais precisamente, para os indivíduos entre 1-17, houve um aumento de 2,06 pontos percentuais em relação à agregação de menores de idade do experimento anterior. Em relação aos maiores de idade, a diferença representou uma acentuação maior, com acréscimo de 16,64 pontos percentuais.

Tabela 6. Resultados da métrica *f1-score* do Experimento 3.

Classes	Precisão	Recall	f1-score	Nº de amostras
1-17	97,60%	96,06%	96,83%	127
18-acima	96,12%	97,64%	96,88%	127

Pela Figura 12, observa-se a matriz de confusão para a melhor performance de classificação do experimento. Das 127 imagens utilizadas como teste para cada classe, 122 foram corretamente classificadas em menores de idade e 124 em maiores de idade. Observou-se que realizar o treinamento binário direto para a rede neural resultou em uma melhora na taxa de amostras corretamente classificadas (*recall*) nas duas faixas.

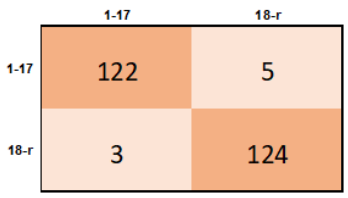

Figura 12. Matriz de confusão do Experimento 3.

6. Conclusões

Este trabalho teve como objetivo central agregar conhecimento à análise pericial, utilizando métodos supervisionados de *deep learning* para realizar estimação de idade. Para os experimentos, os resultados foram próximos aos trabalhos correlatos apresentados na Seção 3. Como o Experimento 1 buscou dividir as classes em um número próximo aos trabalhos existentes, este projeto superou em 3.51 pontos percentuais o proposto por [Anand et al. 2017] e em 1.22 pontos percentuais o proposto por [Rodríguez et al. 2017], com 62,00%. Apesar dos Experimentos 2 e 3 apresentarem divisões menores de classes, os resultados alcançados foram considerados como satisfatórios, com valores médios de 73% e 96,85% de precisão. Em geral, a rede neural densamente conectada se mostrou com alta capacidade performática em tarefas de classificação etária.

Peritos especializados em crimes relacionados à pornografia infantil dispõe de ferramentas que utilizam o aprendizado profundo como método de reconhecimento. Em termos de aplicabilidade, este trabalho mostrou que o uso de modelos neurais convolucionais pode ser seriamente considerado como mecanismo preciso para a identificação de possíveis menores de idade em imagens digitais, considerando diferentes intervalos de faixa etária como alvo de classificação.

Como possibilidades de trabalhos futuros, direcionar a performance do modelo neural para um problema de regressão e comparar com os resultados obtidos neste trabalho. Além disso, unir o reconhecimento etário à uma métrica de identificação de nudez utilizando processos de segmentação de imagens digitais. Caso essa fusão apresente métricas críveis apoiadas por peritos criminais, os resultados podem ser muito relevantes para o avanço dessa área de estudo.

Referências

Anand, A., Labati, R. D., amd Enrique Munoz, A. G., Piuri, V., and Scotti, F. (2017). Age estimation based on face images and pre-trained convolutional neural networks. *2017 IEEE Symposium Series on Computational Intelligence (SSCI)*.

Brasil (1940). Lei no 2.848, de 7 de Dezembro de 1940. Código Penal.

Brasil (1990). Lei no 8.069, de 13 de Julho de 1990. Dispõe sobre o Estatuto da Criança e do Adolescente e dá outras providências.

Buduma, N. and Lacascio, N. (2017). *Fundamentals of Deep Learning*. O'Reilly Media.

Caiado, F. B. and Caiado, M. (2018). Combate à pornografia infantojuvenil com aperfeiçoamentos na identificação de suspeitos e na detecção de arquivos de interesse.

Goodfellow, I., Bengio, Y., and Courville, A. (2016). *Deep Learning*. MIT Press. `http://www.deeplearningbook.org`.

Hamada, F. M. and Sanchez, C. J. P. (2007). Abuso Sexual Infantil: Normatização, Internet e Pedofilia. *Encontro de Iniciação Científica (ETIC)*, 3(3).

Haykin, S. (2008). *Neural networks and learning machines*. Pearson Education, New Jersey, 3 edition.

Hinton, G. E., Srivastava, N., Krizhevsky, A., Sutskever, I., and Salakhutdinov, R. R. (2012). Improving neural networks by preventing co-adaptation of feature detectors. *CoRR*, abs/1207.0580. cite arxiv:1207.0580.

Hope, T., Resheff, Y. S., and Lieder, I. (2017). *Learning TensorFlow: A Guide to Building Deep Learning Systems*. O'Reilly Media, Inc., 1st edition.

Huang, G., Liu, Z., van der Maaten, L., and Weinberger, K. Q. (2016). Densely connected convolutional networks. *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 2261–2269.

Machado, T. J. X. (2017). Cibercrime e o crime no mundo informático.

Mitchell, T. (1997). *Machine Learning*. McGraw-Hill International Editions. McGraw-Hill.

Patterson, J. and Gibson, A. (2017). *Deep Learning*. O'Reilly Media, Sebastopol.

Ponti, M. A. and da Costa, G. B. P. (2018). Como funciona o deep learning. *CoRR*, abs/1806.07908.

Ramos, D. (2018). Pedofilia. *Revista Pericia Federal*.

Raschka, S. (2015). *Python Machine Learning*. Packt Publishing.

Rodríguez, P., Cucurull, G., Gonfaus, J. M., Roca, F. X., and Gonzlez, J. (2017). Age and gender recognition in the wild with deep attention. *Pattern Recogn.*, 72(C):563–571.

Rothe, R., Timofte, R., and Gool, L. V. (2018). Deep expectation of real and apparent age from a single image without facial landmarks. *Int. J. Comput. Vision*, 126(2-4):144–157.

Silva, R. M., Almeida, T. A., and Yamanaki, A. (2012). Análise de desempenho de redes neurais artificiais para classificação automática de web spam.

Simonyan, K. and Zisserman, A. (2014). Very deep convolutional networks for large-scale image recognition. *CoRR*, abs/1409.1556.

TensorFlow (2018). An open source machine learning framework for everyone.