

Estimação Inteligente de Idade Utilizando *Deep Learning*

Nicoli Pinheiro de Araujo

1. Apresentação e Justificativa

A face humana detém diversos atributos que podem ser utilizados para caracterizar indivíduos. Os traços faciais provêm informações como identidade, expressão facial, gênero, origem étnica e idade de uma pessoa. Quanto a esta última, seres humanos são capazes de determinar a idade de indivíduos entre 20 e 60 anos a partir da face com alta acurácia. Porém, há casos em que a checagem de idade realizada por humanos é inconveniente. Neste contexto, pode-se fazer uso de sistemas de determinação automática de idade, especialmente quando é necessário apenas saber a idade de um indivíduo, sem identificá-lo (FU; GUO; HUANG, 2010).

A estimação de idade consiste de rotular a imagem de uma face automaticamente com um número ou uma classe que corresponda a uma idade ou intervalo de idade do indivíduo portador da face fotografada. Estimadores de idade são particularmente úteis em aplicações em que é necessário ou desejável conhecer a idade do usuário para que seja entregue um produto ou serviço de maneira personalizada. Alguns exemplos incluem sistemas de recomendação de conteúdo, enfermeiras robóticas, propagandas direcionadas, entre outros. Há também situações em que a idade do indivíduo determina se haverá ou não o provimento de produto ou serviço, ou se sua entrada será permitida em algum ambiente. Exemplos destes casos incluem sistemas de checagem de idade na entrada de bares e boates, em máquinas de venda automática que contenham bebidas alcóolicas, e sistemas de controle parental (FU; GUO; HUANG, 2010).

Diante do que foi exposto, esta proposta de projeto de mestrado considera o desenvolvimento de estratégias inteligentes, baseadas na utilização de técnicas de *Deep Learning*, para estimação da idade de indivíduos a partir de fotografias faciais. Embora a estimação de outras características também pudesse ser realizada mediante a análise de fotografias faciais, desde gênero até a presença de doenças, optou-se pela idade por ser um atributo comum a todos os indivíduos, pelo potencial de aplicações, pela existência de bases de dados adequadamente rotuladas com este atributo e pelo menor potencial de infringência das searas privadas dos usuários.

A realização de um projeto de mestrado desta natureza é justificada por várias razões. No contexto da interação entre telespectador e *Smart TV*, um estimador de idade pode ser utilizado para facilitar a coleta de informações que contribuam para melhor experiência de provimento de conteúdo e de configurações personalizadas. Em particular, a estimação de idade dos telespectadores pode ser especialmente empregada na implementação de um controle parental mais eficiente, protegendo crianças e adolescentes de conteúdos inadequados à sua faixa etária.

Quanto ao provimento de propaganda direcionada em vitrines e telões distribuídos em shoppings e centros comerciais, a estimação da idade de frequentadores pode ser utilizada juntamente com outras informações como dias, horários e corredores de pico ou histórico de compras de clientes, a fim de que lojas tenham maior conhecimento sobre seus consumidores, de maneira a oferecer ofertas e produtos mais adequados a cada público. As idades aferidas nos referidos locais de livre circulação também podem ser utilizadas para o direcionamento de propagandas governamentais que tenham público-alvo com idade específica, a exemplo de campanhas de vacinação, matrícula em escolas públicas, alistamento para serviço militar, entre outros.

No contexto do desenvolvimento de uma dissertação de mestrado, esta proposta é justificada pela possibilidade de comparação de desempenho de diferentes abordagens para estimação de idade utilizando *Deep Learning* encontradas na literatura. Este aspecto prevê a reprodutibilidade de resultados previamente descritos, e possibilita a aplicação de técnicas e modelos ainda não documentados para esta aplicação. Os resultados obtidos com tais experimentos podem vir a contribuir para o estado da arte neste ramo do conhecimento.

Para a apresentação desta proposta de projeto de mestrado, este documento está organizado como segue. Na Seção 2 estão descritos os objetivos geral e específicos. A Seção 3 contém uma breve contextualização bibliográfica referente a *Deep Learning* e trabalhos relacionados com a estimação de idade a partir de fotografias faciais. A metodologia a ser aplicada durante a execução da pesquisa encontra-se na Seção 4. Por fim, a Seção 5 detalha o cronograma de atividades.

2. Objetivos

O objetivo geral deste trabalho consiste em aplicar técnicas de *Deep Learning* para estimação de idade de indivíduos a partir de suas respectivas fotografias faciais. Para alcançar esta meta, alguns objetivos específicos precisam ser contemplados, a citar:

1. Formular um referencial teórico sobre redes neurais convolucionais, contemplando seu arcabouço matemático, suas características, principais arquiteturas, métodos de treinamento e teste;
2. Consolidar uma base de dados com exemplos realísticos para treinamento dos modelos, tendo em vista a captura de padrões representativos ao domínio do problema;
3. Identificar tecnologias adequadas para implementação dos estimadores;
4. Propor, treinar e testar diferentes estimadores de idade baseados em redes neurais convolucionais para a tarefa em questão;
5. Avaliar comparativamente os estimadores propostos.

3. Contextualização Bibliográfica

A contextualização bibliográfica para a realização deste trabalho compreende conceitos ligados ao *Deep Learning*. Assim, a Seção 3.1 introduz os conceitos essenciais relativos à área. Os conceitos relativos especificamente às redes neurais convolucionais são descritos na Seção 3.2. Alguns trabalhos relacionados à estimação de idade por meio de fotografias faciais estão descritos na Seção 3.3.

3.1. *Deep Learning*

Deep Learning (DL), também conhecido como Aprendizado Profundo, compreende um conjunto de técnicas de Machine Learning (ML) que podem ser aplicadas em problemas de aprendizado supervisionado e não-supervisionado. A principal característica dos modelos neste domínio é a capacidade de representar e reconhecer características sucessivamente complexas, por meio da adição de níveis ou camadas de operações não-lineares em sua arquiteturas, a exemplo das redes neurais profundas, máquinas de Boltzmann profundas e fórmulas proposicionais. Modelos deste tipo ganharam popularidade ao se mostrarem capazes de resolver problemas complexos com um desempenho cada vez maior (BENGIO et al., 2009).

A melhoria do desempenho de modelos de DL é decorrente do aumento recente da quantidade de dados disponíveis sobre temas complexos, aliado ao aumento da disponibilidade de recursos computacionais para executar modelos mais robustos (GOODFELLOW; BENGIO; COURVILLE, 2016; DENG; YU et al., 2014). Alguns dados fornecidos pela IBM reforçam esta afirmação: em 2017 foram gerados 2,5 quintilhões de bytes de dados por dia, e 90% do volume total de dados gerados até 2017 no mundo foi criado somente nos últimos dois anos (IBM, 2017). Estes fatores possibilitaram a implementação de modelos que apresentaram uma melhoria significativa na eficiência de generalização frente a modelos existentes até então, especialmente em virtude da capacidade de organizar a computação como uma composição de várias operações não-lineares (funções de ativação) e uma hierarquia de características re-utilizadas (adição de camadas) (GOODFELLOW; BENGIO; COURVILLE, 2016).

3.2. Redes Neurais Convolucionais

Redes Neurais Convolucionais (CNNs, do inglês *Convolutional Neural Networks*) são uma classe de redes neurais *feedforward* com topologia bem definida e estrutura em grade, com o uso de operações de convolução em pelo menos uma de suas camadas (GOODFELLOW; BENGIO; COURVILLE, 2016). Aplicadas em tarefas de classificação, regressão, localização, detecção e outras, este tipo de modelo se destaca no reconhecimento de padrões em dados de alta dimensionalidade, a exemplo de séries temporais, imagens e vídeos (KHAN et al., 2018).

A operação de convolução possui um papel central nas CNNs. Esta operação descreve a média ponderada de uma determinada função $x_1(t)$ sob um intervalo fixo de uma variável, enquanto os pesos da média ponderada considerada pertencem à função $x_2(t)$ amostrados em intervalos a (BRACEWELL, 1978). Assim, a convolução $s(t)$ de duas funções $x_1(t)$ e $x_2(t)$ é uma função $s : \mathbb{Z} \rightarrow \mathbb{R}$, denotada $s(t) = x_1(t) * x_2(t)$, e definida conforme Equação 1 (LATHI, 2007):

$$s(t) = x_1(t) * x_2(t) = \int_{-\infty}^{\infty} x_1(a)x_2(t-a)da. \quad (1)$$

No contexto de ML, a função $x_1(t)$ é chamada de *input*, a função $x_2(t)$ é o *kernel*, e a saída $s(t)$ consiste no *feature map*, ou mapa de características. No contexto prático, o *input* normalmente é um vetor multidimensional de dados e o *kernel* é um vetor multidimensional de pesos que devem ser ajustados durante o aprendizado das CNNs a fim de mapear um conjunto de entrada a um universo de saída. Considerando, por exemplo, uma imagem I de dimensões (m, n) como *input* e a aplicação de um *kernel* K , a versão discreta da convolução, passível de implementação computacional e equivalente à Equação 1, é mostrada na Equação 2:

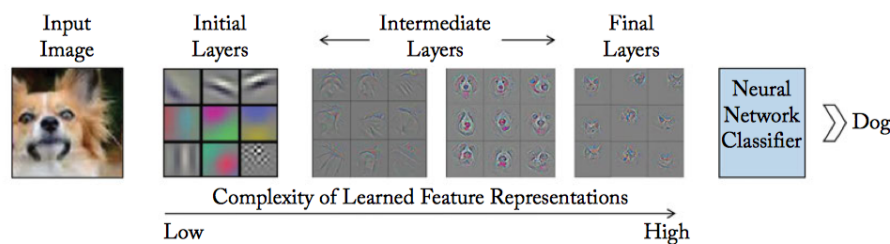
$$S(i, j) = I(i, j) * K(i, j) = \sum_m \sum_n I(m, n)K(i-m, j-n), \quad (2)$$

em que S é o *feature map* resultante e (i, j) é a posição correspondente nesse mapa. Para otimizar os aspectos de implementação, os valores resultantes da operação de convolução são armazenados apenas nas posições (i, j) explicitamente declaradas (GOODFELLOW; BENGIO; COURVILLE, 2016).

Os *feature maps*, resultantes das operações de convolução, compreendem a noção de filtros, responsáveis por capturarem características relativas à entrada, tais como

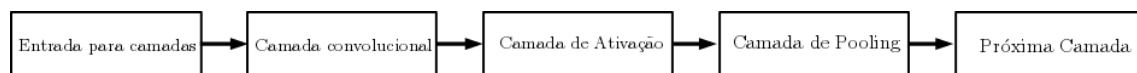
contornos, linhas, texturas, etc. Quando combinados de maneira sequencial, como proposto pelas CNNs, as características capturadas pelas camadas convolucionais se tornam gradativamente mais complexas à medida que a profundidade da rede é incrementada, conforme ilustrado na Figura 1. Assim, um primeiro *feature map* de uma camada convolucional captura um simples contorno, enquanto um *feature map* em uma camada mais profunda da rede pode capturar uma forma, um rosto ou até um objeto inteiro (BUDUMA, 2017).

Figura 1: Papel das camadas convolucionais e *feature maps* nas CNNs. Fonte: (KHAN et al., 2018).



As camadas convolucionais, que capturam os *feature maps* e contém os pesos da rede, normalmente são seguidas por funções de ativação. Via de regra, a toda camada convolucional em uma CNN, segue-se uma função de ativação, finalizando em uma operação de *pooling*, como mostra a Figura 2.

Figura 2: Componentes de uma camada de uma rede neural convolucional. Adaptado de (GOODFELLOW; BENGIO; COURVILLE, 2016).



Uma função de *pooling*, por sua vez, substitui a saída da rede em determinada localização por uma síntese estatística das saídas vizinhas. Por exemplo, a operação *max pooling* retorna o valor máximo em uma área retangular, enquanto a *average pooling* retorna a média das saídas de um retângulo. O objetivo desta operação é fazer com que o *feature map* seja invariante a pequenas mudanças na entrada. Esta invariância a pequenas mudanças locais é uma propriedade útil quando o mais importante for a existência da característica e não exatamente a sua posição. A operação de *pooling* aumenta a eficiência geral da CNN ao reduzir drasticamente o número de valores a serem passados entre duas camadas quaisquer (GOODFELLOW; BENGIO; COURVILLE, 2016).

Embora se tenha uma noção clara das camadas individuais e de suas respectivas funções, a combinação das mesmas em uma rede neural convolucional não é uma tarefa trivial. Este arranjo pode resultar em um número arbitrariamente grande de redes com milhares de parâmetros ajustáveis, cujo desempenho acerca de um problema ainda precisará ser aferido. Considerando os esforços computacionais para isto, a maioria das soluções atuais baseadas em DL fazem uso de CNNs canônicas já propostas na literatura. Alguns trabalhos relacionados com a estimação de idade a partir de fotografias faciais utilizando CNNs estão expostos a seguir.

3.3. Trabalhos Relacionados

A proposta apresentada está relacionada com inúmeros trabalhos que consistem na aplicação de redes neurais convolucionais e outros modelos de ML para a estimação de idade de indivíduos a partir de fotos.

Segundo Fu et al., a idade pode ser inferida a partir de padrões distintos que emergem através da aparência da face (FU; GUO; HUANG, 2010). Técnicas comuns para a estimação da idade envolvem a dedução de modelos matemáticos a partir do estudo do crescimento de medidas da face e do crânio (KWON; LOBO, 1999), da textura do rosto (LANITIS; TAYLOR; COOTES, 2002), da captura de tendências de envelhecimento a partir de várias imagens de indivíduos de mesma idade (FU; XU; HUANG, 2007) e da extração de características específicas relacionadas à idade (SUO et al., 2008; LOU et al., 2018). Modelos de ML também são utilizados para a tarefa, em especial as redes neurais artificiais, K-vizinhos mais próximos e máquinas de vetores de suporte.

Recentemente, a aplicação de redes neurais convolucionais em problemas de classificação e detecção de objetos em imagens têm obtido resultados significativamente positivos. Em diversos trabalhos da literatura são descritas arquiteturas robustas capazes de detectar dezenas de objetos em várias situações (SIMONYAN; ZISSERMAN, 2014; HE et al., 2015; SZEGEDY et al., 2015; REDMON et al., 2015; LIU et al., 2015a). Treinadas com conjuntos de dados visuais que contam com milhares de exemplos como a ImageNet (IMAGENET, 2015), Pascal VOC (EVERINGHAM et al., 2014) e COCO (LIN et al., 2014), estas redes são conhecidas por seu bom desempenho. Algumas destas redes foram ajustadas utilizando conjuntos de dados menores e especializados para a tarefa de estimação de idade.

O trabalho de Rothe em (ROTHER; TIMOFTE; GOOL, 2015) relata um método para estimação de idade aparente em imagens de faces imóveis utilizando DL. O método proposto consiste em detectar uma face em uma imagem, para, em seguida estimar, sua idade. Para esta última tarefa, propôs um conjunto de 20 redes neurais convolucionais classificadoras com arquiteturas VGG-16 pré-treinadas com a base de dados ImageNet, e ajustadas utilizando imagens disponibilizadas pelos sites do IMDb, da Wikipedia, e o conjunto de dados *Looking At People* para anotação de idade aparente. Cada CNN tem como saída um número discreto entre 0 e 100, representando a idade prevista. A saída final do modelo consiste na localização do rosto e na média entre as idades previstas pelas 20 redes para o rosto detectado. A solução atingiu um MAE (*Mean Average Error*) de 3.221 na fase de testes.

Em Liu et al. cria-se um estimador de idade composto pela fusão de um modelo regressor e outro classificador (LIU et al., 2015b). Realiza-se um pré-processamento das imagens de entrada, que envolve a detecção das faces presentes em cada imagem, seguida pela etapa de localização de pontos de referência, como olhos, nariz e boca, e por fim há a normalização das faces. Dois métodos de normalização de face são testados, a normalização exterior e interior. Após este pré-processamento, as imagens resultantes são alimentadas a modelos de CNNs inspiradas na *GoogLeNet* (SZEGEDY et al., 2015). O modelo sofreu modificações em sua arquitetura, como adição de normalização do *batch*, remoção de camadas de *dropout* e perda. Foram treinados e testados modelos cujas entradas sofreram diversas variações no tipo de normalização da face, tamanho do corte dos rostos, tipo de tarefa preditiva, etc. Os modelos resultantes destas variações foram unidos em um conjunto, que conseguiu prever idades com MAE de 3.3345 (LIU et al., 2015b).

Ademais, é possível encontrar resultados satisfatórios para a tarefa de aprendizado

proposta utilizando modelos menos complexos. Com o objetivo de consolidar um método de classificação de idade e gênero, o trabalho de Levi e Hassner propõe uma CNN de natureza mais simples, se comparada com as citadas acima (LEVI; HASSNER, 2015). A arquitetura proposta consiste em três camadas convolucionais com *dropout* e funções de ativação *ReLU*, seguidas por três camadas totalmente conectadas. A camada de saída tem como função de ativação a *Softmax*. A escolha por um design de rede menor é motivado pelo desejo de reduzir o risco de *overfitting* e pela natureza do problema, que contém somente 8 classes de idade. O modelo é treinado utilizando apenas o conjunto de referência *Adience*, composto por imagens não filtradas para classificação de idade e gênero. Considerando uma margem de erro de uma classe de idade vizinha, a melhor rede obteve acurácia de $84.7\% \pm 2.2$ ao empregar uma técnica de sobre-amostragem.

4. Metodologia

A metodologia para o desenvolvimento deste trabalho consiste na realização da *contextualização bibliográfica sobre Deep Learning*, em especial contemplando os conceitos relativos às redes neurais convolucionais. Para tanto, considera-se a literatura desta área para que haja o entendimento das bases matemáticas deste modelo computacional, como funcionam, quais as características e as arquiteturas mais importantes. Neste estudo, além dos aspectos teóricos, serão considerados os ambientes de desenvolvimento, bibliotecas e outras tecnologias para implementação dos conceitos contemplados.

Os demais passos que compõem a metodologia deste trabalho baseiam-se no *fluxo de atividades de machine learning* (MARSLAND, 2009). Inicialmente, haverá a aquisição e o pré-processamento de imagens para *consolidar uma base de dados* para esta tarefa de aprendizado. Nesta etapa, será considerada a literatura e uma base de dados já disponível e apropriadamente anotada, com licença livre de utilização.

A seguir, haverá a *proposição de diferentes modelos de redes neurais convolucionais* para a tarefa de aprendizado considerada. Nesta etapa, serão elencados diferentes parâmetros e hiperparâmetros de configuração, bem como arquiteturas. Estes procedimentos visam consolidar um espaço de busca de modelos que possam endereçar a tarefa de maneira mais eficiente.

O próximo estágio consiste no *treinamento das redes neurais convolucionais* para o problema em questão. Durante este processo, uma parte da base de dados será apresentada aos modelos para que haja o ajuste de pesos, compreendendo o aprendizado das características relevantes. O treinamento das redes ocorrerá utilizando computação em nuvem e computadores disponíveis no Laboratório de Sistemas Inteligentes (LSI), tendo em vista a infra-estrutura de hardware necessária para realizar este procedimento.

Segue-se então o *teste das redes*, respeitando uma abordagem de validação cruzada e utilizando métricas de desempenho apropriadas. O objetivo desta fase consiste em aferir os modelos propostos e treinados quanto à sua capacidade de generalização.

Por fim, para identificação de um modelo mais adequado à esta tarefa, as *métricas de desempenho serão comparadas* e os melhores modelos elencados a partir destes valores, apontando assim um estimador apropriado para o problema inicialmente considerado.

Além destas atividades, há que se considerar a escrita e defesa do projeto final de dissertação de mestrado.

5. Cronograma

O cronograma de realização das atividades pode ser visto na Tabela 1. As atividades listadas possuem relação com a metodologia detalhada na seção anterior, compreendendo os requisitos elementares para a realização deste trabalho.

Tabela 1: Cronograma de atividades levando em consideração os vinte e quatro meses (de 04/2019 a 04/2021) para a realização do trabalho final.

		2019												2020												2021			
		04	05	06	07	08	09	10	11	12	01	02	03	04	05	06	07	08	09	10	11	12	01	02	03	04			
Escrita da Dissertação					X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X			
Contextualização Bibliográfica sobre CNNs	X	X	X	X	X	X	X	X	X	X	X	X	X	X															
Consolidação da Base de Dados						X	X	X	X	X	X	X	X	X	X	X													
Proposição de Modelos de CNNs										X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X				
Treinamento das CNNs										X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X				
Teste das CNNs										X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X				
Comparação de Métricas de Desempenho											X	X	X	X	X	X	X	X	X	X	X	X	X	X	X				
Defesa da Dissertação																									X				

Referências

- BENGIO, Y. et al. Learning deep architectures for ai. *Foundations and trends® in Machine Learning*, Now Publishers, Inc., v. 2, n. 1, p. 1–127, 2009.
- BRACEWELL, R. *The Fourier Transform and its Applications*. Second. Tokyo: McGraw-Hill Kogakusha, Ltd., 1978.
- BUDUMA, N. *Fundamentals of Deep Learning*. Estados Unidos: Editora O'Reilly, 2017.
- DENG, L.; YU, D. et al. Deep learning: methods and applications. *Foundations and Trends® in Signal Processing*, Now Publishers, Inc., v. 7, n. 3–4, p. 197–387, 2014.
- EVERINGHAM, M. et al. *The PASCAL Visual Object Classes Homepage*. The PASCAL VOC project, 2014. Acessado em 13 de Junho de 2018. Disponível em: <<http://host.robots.ox.ac.uk/pascal/VOC/>>.
- FU, Y.; GUO, G.; HUANG, T. S. Age synthesis and estimation via faces: A survey. *IEEE transactions on pattern analysis and machine intelligence*, IEEE, v. 32, n. 11, p. 1955–1976, 2010.
- FU, Y.; XU, Y.; HUANG, T. S. Estimating human age by manifold analysis of face pictures and regression on aging features. In: IEEE. *2007 IEEE International Conference on Multimedia and Expo*. Beijing, China, 2007. p. 1383–1386. ISSN 1945-7871.
- GOODFELLOW, I.; BENGIO, Y.; COURVILLE, A. *Deep Learning*. MIT Press, 2016. Disponível em: <<http://www.deeplearningbook.org>>.
- HE, K. et al. Deep residual learning for image recognition. *CoRR*, abs/1512.03385, 2015. Disponível em: <<http://arxiv.org/abs/1512.03385>>.
- IBM, M. C. *10 Key Marketing Trends for 2017 and Ideas for Exceeding Customer Expectations*. 2017. Acessado em 23 de Março de 2018. Disponível em: <<https://www-01.ibm.com/common/ssi/cgi-bin/ssialias?htmlfid=WRL12345USEN>>.

IMAGENET. *Large Scale Visual Recognition Challenge (ILSVRC)*. 2015. Acessado em 31 de Maio de 2018. Disponível em: <<http://image-net.org/challenges/LSVRC/>>.

KHAN, S. et al. *A Guide to Convolutional Neural Networks for Computer Vision*. Austrália: Morgan & Claypool, 2018.

KWON, Y. H.; LOBO, N. da V. Age classification from facial images. *Computer vision and image understanding*, Elsevier, v. 74, n. 1, p. 1–21, 1999.

LANITIS, A.; TAYLOR, C. J.; COOTES, T. F. Toward automatic simulation of aging effects on face images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, IEEE, v. 24, n. 4, p. 442–455, 2002.

LATHI, B. *Sinais e Sistemas Lineares - 2.ed.* BOOKMAN COMPANHIA ED, 2007. ISBN 9788560031139. Disponível em: <<https://books.google.com.br/books?id=ySxoo2TVeeYC>>.

LEVI, G.; HASSNER, T. Age and gender classification using convolutional neural networks. In: *IEEE Conf. on Computer Vision and Pattern Recognition (CVPR) workshops*. IEEE, 2015. Disponível em: <https://www.openu.ac.il/home/hassner/projects/cnn_agegender>.

LIN, T. et al. Microsoft COCO: common objects in context. *CoRR*, abs/1405.0312, 2014. Disponível em: <<http://arxiv.org/abs/1405.0312>>.

LIU, W. et al. Ssd: Single shot multibox detector. In: . Amsterdam, The Netherlands: CoRR, 2015. abs/1512.02325. Disponível em: <<http://arxiv.org/abs/1512.02325>>.

LIU, X. et al. Agetnet: Deeply learned regressor and classifier for robust apparent age estimation. In: *Proceedings of the IEEE International Conference on Computer Vision Workshops*. Santiago, Chile: IEEE, 2015. p. 16–24.

LOU, Z. et al. Expression-invariant age estimation using structured learning. *IEEE transactions on pattern analysis and machine intelligence*, IEEE, v. 40, n. 2, p. 365–375, 2018.

MARSLAND, S. *Machine Learning: An Algorithmic Perspective*. 1st. ed. London, UK: Chapman & Hall/CRC, 2009. ISBN 1420067184, 9781420067187.

REDMON, J. et al. You only look once: Unified, real-time object detection. In: . CoRR, 2015. abs/1506.02640. Disponível em: <<http://arxiv.org/abs/1506.02640>>.

ROTHER, R.; TIMOFTE, R.; GOOL, L. V. Dex: Deep expectation of apparent age from a single image. In: *2015 IEEE International Conference on Computer Vision Workshop (ICCVW)*. Santiago, Chile: IEEE, 2015. v. 00, p. 252–257. Disponível em: <doi.ieeecomputersociety.org/10.1109/ICCVW.2015.41>.

SIMONYAN, K.; ZISSERMAN, A. Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*, 2014.

SUO, J. et al. Design sparse features for age estimation using hierarchical face model. In: *Automatic Face & Gesture Recognition, 2008. FG'08. 8th IEEE International Conference on*. Amsterdam, The Netherlands: IEEE, 2008. p. 1–6.

SZEGEDY, C. et al. Going deeper with convolutions. In: *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. Boston, MA, USA: IEEE, 2015. p. 1–9. ISSN 1063-6919.