



UFRR

UNIVERSIDADE FEDERAL DE RORAIMA
DEPARTAMENTO DE CIÊNCIA DA COMPUTAÇÃO
BACHARELADO EM CIÊNCIA DA COMPUTAÇÃO

**AGE-GAN: Um sistema para geração de imagens
baseado em previsão temporal para pessoas
desaparecidas**

João Paulo Parreira Peixoto

Boa Vista - RR

Outubro de 2021

João Paulo Parreira Peixoto

AGE-GAN: Um sistema para geração de imagens
baseado em previsão temporal para pessoas
desaparecidas

Monografia de Graduação apresentada ao Departamento de Ciência da Computação da Universidade Federal de Roraima como requisito parcial para a obtenção do grau de bacharel em Ciência da Computação.

Orientador(a)

Professor Doutor Herbert Oliveira Rocha

Universidade Federal de Roraima
Departamento de Ciência da Computação

Boa Vista - RR

Outubro de 2021

Dados Internacionais de Catalogação na publicação (CIP)
Biblioteca Central da Universidade Federal de Roraima

P379a Peixoto, João Paulo Parreira.

AGE-GAN: um sistema para geração de imagens baseado em
previsão temporal para pessoas desaparecidas / João Paulo Parreira
Peixoto. – Boa Vista, 2021.

155 f. : il.

Orientador(a) Professor Doutor Herbert Oliveira Rocha.

Monografia (graduação) – Universidade Federal de Roraima,
Curso de Ciência da Computação.

1 – GANs. 2 – Transferência de Estilo. 3 – StyleGAN. 4 –
Previsão Temporal. 5 – Espaço Latente. I – Título. II – Rocha,
Herbert Oliveira (orientador).

CDU – 681:34

DECLARAÇÃO DE AUTORIA

Eu, **João Paulo Parreira peixoto** (código de matrícula **1201524425**), autor da(o) monografia/TCC (Trabalho de Conclusão de Curso) sob o título **AGE-GAN: Um sistema para geração de imagens baseado em previsão temporal para pessoas desaparecidas**, declaro que o trabalho em referência é de minha total autoria e de minha inteira responsabilidade o texto apresentado. Declaro, ainda, que as citações e paráfrases dos autores estão indicadas com as respectivas obras e anos de publicação. Declaro, para os devidos fins que estou ciente:

- dos Artigos 297 a 299 do Código Penal, Decreto-Lei n. 2.848 de 7 de dezembro de 1940;
- da Lei n. 9.610, de 19 de fevereiro de 1998, sobre os Direitos Autorais; e
- que plágio consiste na reprodução de obra alheia e submissão da mesma como trabalho próprio ou na inclusão, em trabalho próprio, de ideias, textos, tabelas ou ilustrações (quadros, figuras, gráficos, fotografias, retratos, lâminas, desenhos, organogramas, fluxogramas, plantas, mapas e outros) transcritos de obras de terceiros sem a devida e correta citação da referência.

O corpo docente responsável pela avaliação deste trabalho poderá não aceitar o referido trabalho caso os pontos mencionados acima sejam descumpridos, por conseguinte, considerar-me reprovado.

Assinatura do acadêmico(a)
Boa Vista - RR, 22 de Outubro de 2021.

FOLHA DE APROVAÇÃO

Monografia de Graduação sob o título **AGE-GAN: Um sistema para geração de imagens baseado em previsão temporal para pessoas desaparecidas** apresentada por **João Paulo Parreira Peixoto** e aceita pelo Departamento de Ciência da Computação da Universidade Federal de Roraima, sendo aprovada por todos os membros da banca examinadora abaixo especificada:

Prof. Dr. Herbert Oliveira Rocha
Orientador(a)
Departamento de Ciência da Computação
Universidade Federal de Roraima

Prof. Dr. Felipe Leite Lobo
Departamento de Ciência da Computação
Universidade Federal de Roraima

Prof. Dr. Leandro Nelinho Balico
Departamento de Ciência da Computação
Universidade Federal de Roraima



Boa Vista - RR, 22 de Outubro de 2021.

Dedico este trabalho a minha mãe, pai, irmão, tios e padrinho, que foram de grande inspiração em minha jornada acadêmica e pessoal; que me apoiaram e me incentivaram em momentos difíceis e decisivos da minha vida.

Agradecimentos

Agradeço primeiramente a minha mãe Janete Aparecida Parreira, que sempre fez tudo o que estava ao seu alcance e até mesmo o que não estava em alguns momentos para mim e meu irmão. Também agradeço a meu pai João Martins Peixoto Junior, que esteve me aconselhando, apoiando e incentivando ao longo dos anos. Agradeço ao meu irmão Alisson Junio Parreira Peixoto, por toda a atenção, apoio e incentivo me dado ao longo dos anos. Neste agradecimento também incluo meus tios, Núbia Maria Parreira e Fábio José Parreira, que foram meus segundos pais, sempre apoiando e incentivando.

Ao meu padrinho Luciano Ferreira Silva, que foi meu terceiro pai, ao longo da vida sempre aconselhando, apoiando e incentivando em todos os momentos importantes.

A todos os professores que ao longo da minha vida acadêmica contribuíram de alguma forma para que eu pudesse chegar no ponto em que estou hoje. Agradeço principalmente aos meus professores da universidade por me guiar e me orientar nessa longa jornada da universidade. Também gostaria de agradecer ao Professor Msc. Filipe Dwan Pereira, por todos os conselhos dados, pela paciência e por seus ensinamentos durante as disciplinas.

Agradeço especialmente ao meu orientador Professor Dr. Herbert Oliveira Rocha, por ter me guiado ao longo do caminho universitário, pelos ensinamentos tanto nas disciplinas quanto principalmente durante a realização do trabalho de pesquisa e também por aconselhar em momentos de desespero.

Por fim, mas não menos importante, gostaria de agradecer a todas as outras pessoas que de alguma forma contribuíram para o meu crescimento tanto pessoal quanto acadêmico.

Sempre serei grato a todos que passaram ou ainda fazem parte da minha vida e contribuíram de alguma forma no pessoal e/ou acadêmico até mesmo nos momentos em que nem eu mesmo acreditava em mim e/ou não era uma pessoa fácil de lidar.

A persistência é o caminho do êxito.

Charles Chaplin

AGE-GAN: Um sistema para geração de imagens baseado em previsão temporal para pessoas desaparecidas

Autor: João Paulo Parreira Peixoto

Orientador: Professor Doutor Herbert Oliveira Rocha

Resumo

Recentemente o número crescente de casos de pessoas desaparecidas tem sido um problema global, necessitando urgentemente de atenção. Este trabalho visa buscar uma solução para minimizar o problema, propondo e avaliando ferramentas para auxiliar nas investigações de desaparecidos. As redes adversárias generativas, do inglês *Generative Adversarial Networks (GANs)* que são suportadas por redes neurais artificiais apresentam resultados significativos na geração de imagens. Assim, este trabalho propõe a utilização de um modelo baseado em GANs que se baseiam na transferência de estilo (StyleGAN) com o propósito para a geração de imagens baseado em uma previsão temporal do rosto de pessoas desaparecidas. Os resultados obtidos na avaliação experimental deste trabalho demonstraram que o modelo possui melhores resultados ao se trabalhar no espaço latente W , e que a aplicação cumpre o que promete, possuindo de formas gerais resultados promissores. No entanto, foi percebido que a aplicação também erra bastante em configurações específicas (exemplo, cabelos longos), sendo estes pontos trabalhos futuros a serem implementados.

Palavras-chave: GANs, Transferência de Estilo, StyleGAN, Previsão Temporal, Espaço Latente.

AGE-GAN: Um sistema para geração de imagens baseado em previsão temporal para pessoas desaparecidas

Autor: João Paulo Parreira Peixoto

Orientador: Professor Doutor Herbert Oliveira Rocha

Abstract

Recently, the growing number of missing persons cases has been a global problem, needing urgent attention. This work aims to find a solution to minimize the problem, proposing and evaluating tools to assist in the investigation of missing persons. The generative adversarial networks (GANs) which are supported by artificial neural networks, present significant results in the generation of images. Therefore, this work proposes the use of a model based on GANs based on style transfer (StyleGAN) with the purpose to generate images based on a temporal forecast of the face of missing persons. The results obtained in the experimental evaluation of this work showed that the model has better results when working in the latent space W , and that the application fulfills what it promises, having generally promising results. However, it was noticed that the application also misses a lot in specific settings (eg, long hair), and these points are future works to be implemented.

Keywords: GANs, Style Transfer, StyleGAN, Temporal Forecast, Latent Space.

Lista de figuras

Figura 1 – Exemplos de fotografias geradas com o uso de GANs.	19
Figura 2 – Exemplos de fotografias geradas com o uso de GANs.	20
Figura 3 – Ramos da Inteligência Artificial.	25
Figura 4 – Exemplo de algoritmos para cada tipo de aprendizado.	26
Figura 5 – Fluxo de aprendizagem supervisionada.	27
Figura 6 – Exemplo de estrutura <i>single-layer perceptron</i>	29
Figura 7 – Exemplo de estrutura <i>multi-layer feed-foward</i> totalmente conectada.	30
Figura 8 – Estrutura de GANs.	32
Figura 9 – Taxonomia proposta pelos autores.	34
Figura 10 – Ilustração de uma estrutura simples de GAN condicional. . .	36
Figura 11 – Arquitetura da estrutura expandida de uma cGAN.	37
Figura 12 – Estrutura de cGANs.	38
Figura 13 – Arquitetura do Age-cGAN proposto pela equipe.	42
Figura 14 – Arquitetura do modelo de <i>GAN-Wavelet</i> proposto pela equipe.	44
Figura 15 – Manipulações de vários atributos faciais por meio da variação dos códigos latentes de um modelo de GAN bem treinado. . .	47
Figura 16 – Amostras de síntese com a distância próxima (linha do meio) e extremamente longe (linha superior e inferior) do limite de separação, por atributo.	49
Figura 17 – Análise do espaço latente Z e do espaço latente desemaranhado W do StyleGAN tomando a manipulação da idade como exemplo.	50
Figura 18 – Arquitetura de um gerador tradicional (a), e arquitetura do gerador baseado em estilo proposto pelos autores (b).	52
Figura 19 – Funcionamento do sistema proposto.	59
Figura 20 – Diagrama de fluxo.	60
Figura 21 – Estrutura interna de um vetor latente.	65

Figura 44 – Resultados para a qualidade da previsão de envelhecimento do modelo gerador.	89
Figura 45 – Resultados para a qualidade da previsão de envelhecimento do modelo gerador por celebridade.	90
Figura 46 – Resultados do nível de satisfação com a aplicação	91
Figura 47 – Início do questionário - parte 1	99
Figura 48 – Início do questionário - parte 2	100
Figura 49 – Qualidade das imagens - parte 1	100
Figura 50 – Qualidade das imagens - parte 2	101
Figura 51 – Qualidade das imagens - parte 3	102
Figura 52 – Qualidade das imagens - parte 4	103
Figura 53 – Qualidade das imagens - parte 5	104
Figura 54 – Características Faciais - Will Smith - parte 1	105
Figura 55 – Características Faciais - Will Smith - parte 2	106
Figura 56 – Características Faciais - Will Smith - parte 3	107
Figura 57 – Características Faciais - Will Smith - parte 4	108
Figura 58 – Características Faciais - Will Smith - parte 5	109
Figura 59 – Características Faciais - Eunice Baía - parte 1	110
Figura 60 – Características Faciais - Eunice Baía - parte 2	111
Figura 61 – Características Faciais - Eunice Baía - parte 3	112
Figura 62 – Características Faciais - Eunice Baía - parte 4	112
Figura 63 – Características Faciais - Eunice Baía - parte 5	113
Figura 64 – Características Faciais - Daniele Suzuki - parte 1	114
Figura 65 – Características Faciais - Daniele Suzuki - parte 2	115
Figura 66 – Características Faciais - Daniele Suzuki - parte 3	116
Figura 67 – Características Faciais - Daniele Suzuki - parte 4	116
Figura 68 – Características Faciais - Daniele Suzuki - parte 5	117
Figura 69 – Características Faciais - Leonardo Dicaprio - parte 1	118
Figura 70 – Características Faciais - Leonardo Dicaprio - parte 2	119
Figura 71 – Características Faciais - Leonardo Dicaprio - parte 3	120

Figura 72 – Características Faciais - Leonardo Dicaprio - parte 4	121
Figura 73 – Características Faciais - Leonardo Dicaprio - parte 5	121
Figura 74 – Qualidade do envelhecimento - parte 1	122
Figura 75 – Qualidade do envelhecimento - parte 2	123
Figura 76 – Qualidade do envelhecimento - parte 3	124
Figura 77 – Grau de conhecimento/Expertise - parte 1	125
Figura 78 – Grau de conhecimento/Expertise - parte 2	125
Figura 79 – Satisfação	126
Figura 80 – Resultados - Início do questionário	127
Figura 81 – Resultados - Qualidade das imagens - parte 1	128
Figura 82 – Resultados - Qualidade das imagens - parte 2	128
Figura 83 – Resultados - Características Faciais - Will Smith - parte 1 . .	129
Figura 84 – Resultados - Características Faciais - Will Smith - parte 2 . .	130
Figura 85 – Resultados - Características Faciais - Will Smith - parte 3 . .	131
Figura 86 – Resultados - Características Faciais - Will Smith - parte 4 . .	131
Figura 87 – Resultados - Características Faciais - Will Smith - parte 5 . .	132
Figura 88 – Resultados - Características Faciais - Will Smith - parte 6 . .	133
Figura 89 – Resultados - Características Faciais - Will Smith - parte 7 . .	134
Figura 90 – Resultados - Características Faciais - Eunice Baía - parte 1 .	135
Figura 91 – Resultados - Características Faciais - Eunice Baía - parte 2 .	136
Figura 92 – Resultados - Características Faciais - Eunice Baía - parte 3 .	137
Figura 93 – Resultados - Características Faciais - Eunice Baía - parte 4 .	137
Figura 94 – Resultados - Características Faciais - Eunice Baía - parte 5 .	138
Figura 95 – Resultados - Características Faciais - Eunice Baía - parte 6 .	139
Figura 96 – Resultados - Características Faciais - Daniele Suzuki - parte 1	140
Figura 97 – Resultados - Características Faciais - Daniele Suzuki - parte 2	141
Figura 98 – Resultados - Características Faciais - Daniele Suzuki - parte 3	142
Figura 99 – Resultados - Características Faciais - Daniele Suzuki - parte 4	142
Figura 100–Resultados - Características Faciais - Daniele Suzuki - parte 5	143
Figura 101–Resultados - Características Faciais - Daniele Suzuki - parte 6	144

Figura 102–Resultados - Características Faciais - Daniele Suzuki - parte 7	145
Figura 103–Resultados - Características Faciais - Leonardo Dicaprio - parte 1	146
Figura 104–Resultados - Características Faciais - Leonardo Dicaprio - parte 2	147
Figura 105–Resultados - Características Faciais - Leonardo Dicaprio - parte 3	148
Figura 106–Resultados - Características Faciais - Leonardo Dicaprio - parte 4	148
Figura 107–Resultados - Características Faciais - Leonardo Dicaprio - parte 5	149
Figura 108–Resultados - Características Faciais - Leonardo Dicaprio - parte 6	150
Figura 109–Resultados - Qualidade do envelhecimento - parte 1	151
Figura 110–Resultados - Qualidade do envelhecimento - parte 2	152
Figura 111–Resultados - Qualidade do envelhecimento - parte 3	153
Figura 112–Resultados - Grau de conhecimento/Expertise - parte 1	154
Figura 113–Resultados - Grau de conhecimento/Expertise - parte 2	154
Figura 114–Resultados - Satisfação	155

Lista de tabelas

Tabela 1 – Resultados de estimativa de idade encontrados pela equipe em Morph e CACD (diferenças de idades médias são medidas em valor absoluto)	45
Tabela 2 – Resultados encontrados pela equipe na verificação facial em Morph e CACD	46
Tabela 3 – Resultados encontrados pela equipe nas taxas de preservação de atributos faciais para 'Genero' e 'Raça' em Morph e CACD	46
Tabela 4 – Resultados em porcentagem de acurácia de classificação nos limites de separação no espaço latente em relação a diferentes atributos.	48
Tabela 5 – Comprimentos de caminho de percepção e pontuações de separabilidade para várias arquiteturas de gerador em FFHQ (quanto menor, melhor).	54
Tabela 6 – O efeito de uma rede de mapeamento em FFHQ. O número no nome do método indica a profundidade da rede de mapeamento.	54
Tabela 7 – Classificação dos artigos por técnicas.	57
Tabela 8 – Resultados obtidos por meio das métricas PPL e separabilidade linear para cada espaço latente.	79

Sumário

1	INTRODUÇÃO	17
1.1	Motivação	19
1.2	Definição do Problema	20
1.3	Objetivos	21
1.3.1	Objetivo Geral	21
1.3.2	Objetivos Específicos	21
1.4	Organização do Trabalho	22
2	FUNDAMENTOS TEÓRICOS	24
2.1	Inteligência Artificial (IA)	24
2.1.1	Aprendizado de Máquina	26
2.1.2	<i>Deep Learning</i>	28
2.1.3	<i>Generative Adversarial Networks (GANs)</i>	30
2.1.3.1	Taxonomia de modelos generativos	32
2.1.3.2	AutoEncoders	34
2.1.4	GANs Condicionais	35
2.2	<i>Frameworks para implementação de GANs</i>	38
2.3	Destaques conceituais para pesquisa	40
3	TRABALHOS CORRELATOS	41
3.1	Trabalho 1 - <i>Face aging with conditional generative adversarial networks</i>	41
3.2	Trabalho 2 - <i>Attribute-aware Face Aging with Wavelet-based Generative Adversarial Networks</i>	43
3.3	Trabalho 3 - <i>Interpreting the Latent Space of GANs for Semantic Face Editing</i>	46
3.4	Trabalho 4 - <i>A Style-Based Generator Architecture for Generative Adversarial Networks</i>	50

3.5	Correlações entre os trabalhos e a pesquisa	55
4	MÉTODO DA SOLUÇÃO PROPOSTA	58
4.1	Arquitetura do Método Proposto	58
4.2	Ferramentas e Implementações	61
4.3	Estrutura do GAN-AGE e o Fluxo de Execução	61
5	AVALIAÇÃO EXPERIMENTAL	70
5.1	Planejamento e Projeto da Avaliação Experimental	70
5.2	Execução dos experimentos e análise dos resultados	71
5.2.1	Prova de conceito	71
5.2.2	Análise do modelo GAN por meio de métricas	78
5.2.3	Análise do modelo GAN na geração de imagens do rosto humano	81
6	CONSIDERAÇÕES FINAIS E TRABALHOS FUTUROS	92
	Referências	94
	APÊNDICE A – QUESTIONÁRIO	99
A.1	Estrutura do questionário e questões de pesquisa	99
A.2	Resultados coletados da avaliação	126

1 Introdução

Os casos de pessoas desaparecidas é um problema global que vem crescendo a cada ano que se passa, a uma taxa superior a 10% anualmente, de acordo com o site do Agência Senado (VIEIRA, 2018). No Brasil são registrados, em média, 50 mil casos de desaparecimentos de crianças e adolescentes por ano. Estima-se ainda que quase 250 mil estejam desaparecidos no país. Indo mais além, dados apontam a estimativa de 25 milhões de desaparecimentos no mundo segundo a ONU (VIEIRA, 2018). Alguns países no exterior, a situação é ainda mais alarmante do que no Brasil, como os Estados Unidos. Segundo o site globalmissingkids.org, o número de crianças desaparecidas chega a ser estimado em 460 mil todos os anos (GLOBALMISSINGKIDS, 2020).

Diversas causas podem ser o fator para que estes desaparecimentos aconteçam podendo ser oriundos de um desaparecimento comum, ou até então de vários tipos de crimes, como por exemplo tráfico de pessoas, órgãos, sequestro com roubo seguido de morte, dentre outros. Podemos perceber que este assunto abordado pode levar a muitos outros tipos de crimes.

Pensando nessa problemática, a tecnologia surge se mostrando uma grande aliada para muitas tarefas, dentre essas, auxiliar com o reconhecimento destes desaparecidos. Para que isto seja feito, contamos com a área da inteligência artificial e algumas de suas derivações como *machine learning* e *deep learning*, no campo de envelhecimento facial, através de técnicas para este fim, como as AGE-GAN (do inglês AGE - *Generative Adversarial Network*) (LIU et al., 2017). Com o auxílio da devida tecnologia para reconhecimento, poderíamos ajudar a facilitar a tarefa que as autoridades competentes tem de encontrar alguns destes desaparecidos e criminosos foragidos ou não.

O envelhecimento facial, pode ser conhecido de várias formas, e dentre estas estão a síntese etária (ANTIPOV et al., 2017) ou envelhecimento de rosto humano (LIU et al., 2017). Além do mais, se mostrou ser uma tecnologia que vem sendo muito utilizada por diversos pesquisadores, com alguns propósi-

tos específicos, como auxiliar na busca por desaparecidos ou até então no reconhecimento por foragidos.

As AGE-GAN são aplicações de estudo no ramo do envelhecimento facial. Existem muitas utilidades para a mesma, como o uso para geração de rostos de anime (JIN et al., 2017) e geração de objetos 3D (WU et al., 2017a). Normalmente as GANs são mais usadas para prever e envelhecer um determinado rosto humano (progressão de idade), ou rejuvenescer (regressão de idade), e assim, profetizar a idade do rosto gerado. Nessa técnica são utilizadas duas redes adversárias generativas (do inglês *Generative Adversarial Networks* - GANs), sendo elas o gerador que uma fica responsável por gerar os rostos (os dados não originais), e o discriminador por detectar se aquele rosto gerado é verdadeiro ou falso. Podemos fazer uma analogia que em uma GAN, a rede responsável por gerar dados falsos é uma falsificadora (Gerador) com a missão de enganar a outra rede (Discriminador), e a rede do discriminador, é como se fosse a polícia, com a função de desmascarar a sua rede adversária falsificadora (Gerador).

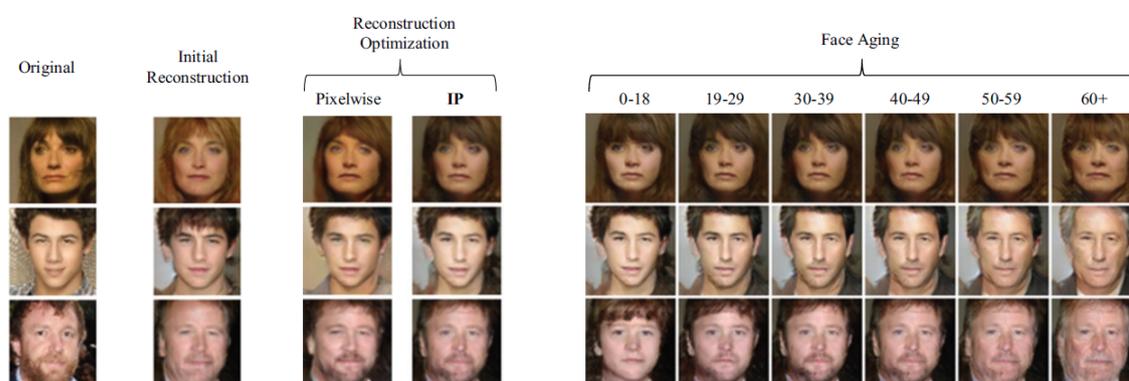
O enfoque deste trabalho é a realização de sínteses de rostos humanos para uma idade futura, ou seja, realizando uma progressão de idade, por meio de GANs, a partir de algumas fotos de um determinado banco de dados, afim de obter uma taxa de precisão aceitável na predição das idades.

No entanto para a síntese das faces, existem alguns desafios para atingir a este fim, como a própria iluminação do ambiente, a posição da face, se a pessoa da foto de entrada possui todas as características faciais bem definidas, e até mesmo tatuagens e objetos que impeçam a síntese, como os bonés e óculos. Além disso, após a foto ser gerada pode existir a dificuldade em realizar o reconhecimento dessa pessoa, pois não sabemos como ela pode estar no futuro, por exemplo, se tem uma nova cicatriz ou uma nova tatuagem.

1.1 Motivação

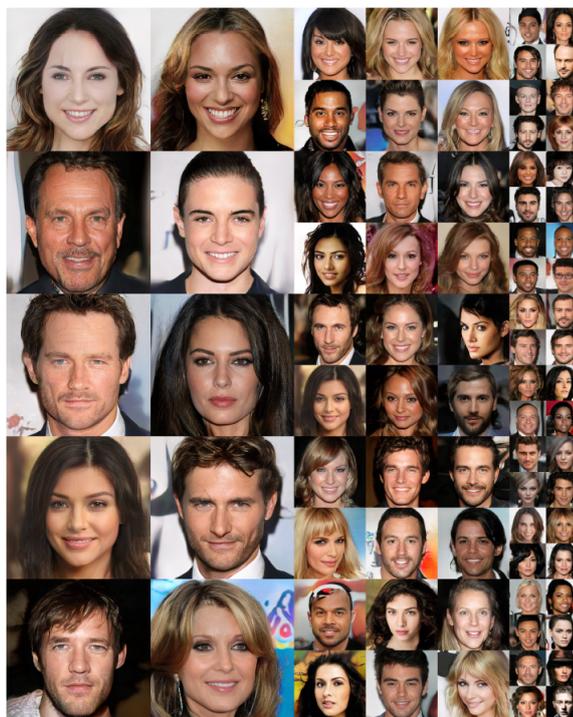
Observando os dados mencionados anteriormente que surge a motivação por realizar uma pesquisa com um sistema para geração de imagens com o foco de previsão temporal da idade futura de pessoas desaparecidas. Outros fatores que podemos pontuar como motivação seriam os trabalhos já realizados utilizando as GANs como meio. Se tratando de aplicações e soluções propostas com temáticas similares ao que pretendemos realizar neste trabalho, podemos citar algumas pesquisas com resultados motivadores. Como no trabalho de (ANTIPOV et al., 2017), onde utilizaram GANs para gerar fotografias de faces com diferentes aparências de idade, como podemos ver na Figura 1; e a pesquisa realizada por (KARRAS et al., 2018), nela é proposto a geração de fotos realísticas de rostos humanos, sendo eles bem similares aos rostos originais, podemos visualizar alguns dos resultados na Figura 2.

Figura 1 – Exemplos de fotografias geradas com o uso de GANs.



Fonte: (ANTIPOV et al., 2017)

Figura 2 – Exemplos de fotografias geradas com o uso de GANs.



Fonte: (KARRAS et al., 2018)

Atualmente as GANs vem sendo cada vez mais utilizadas por grandes empresas e indústrias como a Google que possui sua aplicação de dispositivo móvel capaz de detectar o texto com a câmera do celular e traduzi-lo; A IBM, que possui uma aplicação chamada *IBM Maximo Visual Inspection™*, uma aplicação capaz de monitorar, detectar e solucionar problemas e anomalias (IBM, 2021); e na indústria de saúde, onde as GANs podem ser utilizadas para a detecção de complicações na saúde, como a detecção de tumores (JOSHI, 2020).

1.2 Definição do Problema

A obtenção de uma imagem foto realista futura ou algo semelhante poderia ser uma forma de tornar a investigação e conseqüentemente a localização do paradeiro do desaparecido mais ágil.

Porém, para a geração existem alguns fatores que dificultam o reconhecimento e envelhecimento facial, como a iluminação do ambiente em que a foto foi tirada, posição do rosto, se existem objetos dificultando a detecção do rosto (bonés, *piercings*, tatuagens), ou até então se na foto de um dado banco de imagens o desaparecido não possuía alguma característica facial relevante, mas possui no futuro.

Diante dos dados e fatos apresentados, o problema considerado neste trabalho pode ser expresso com a seguinte questão: **Como uma imagem de um rosto humano pode ser envelhecida, considerando intervalos de idades por meio da utilização de GANs, para auxiliar em investigações de desaparecidos?**

1.3 Objetivos

Esta seção se divide da seguinte forma: objetivos geral e os objetivos específicos. Nas próximas subseções apresentaremos ambos os objetivos que pretendemos manter como enfoque ao longo desta pesquisa.

1.3.1 Objetivo Geral

O objetivo principal deste trabalho é projetar, desenvolver e avaliar um modelo usando GANs para geração de imagens do rosto humano envelhecido baseado em intervalos de idades para uma estimativa na previsão temporal de pessoas desaparecidas, visando assim contribuir em investigações de desaparecidos.

1.3.2 Objetivos Específicos

Os objetivos específicos são:

1. Avaliar os tipos GANs para utilização em sistema gerador de imagens

- preditivo, para determinar as possíveis idades dos desaparecidos;
2. Efetuar um levantamento de banco de imagens utilizados na literatura para o treino de modelos usando GANs;
 3. Desenvolver ou adaptar um modelo de GAN, por meio de técnicas identificadas em publicações acadêmicas e pesquisas já realizadas.
 4. Avaliar a solução proposta por meio da utilização de banco de imagens públicos (para validação do modelo com uso de métricas de aprendizagem de máquina) e também a aplicação de questionário online de pesquisa para o público.

1.4 Organização do Trabalho

A seção atual deste trabalho, capítulo um, apresentou uma breve introdução e noção do problema, motivação e os objetivos que levaram a esta pesquisa. Em relação aos capítulos restantes, eles foram organizados da seguinte forma:

- No Capítulo 2, **Fundamentos Teóricos**, são apresentados os conceitos abordados neste trabalho, e toda a fundamentação teórica pesquisada.
- No Capítulo 3, **Trabalhos Correlatos**, são analisadas e discutidas as partes pertinentes de cada um dos trabalhos correlatos a solução proposta.
- No Capítulo 4, **Método da Solução Proposta**, é descrito as etapas de execução do método da solução proposta.
- No Capítulo 5, **Avaliação Experimental**, é apresentado o planejamento e projeto dos experimentos para a execução da avaliação da solução proposta. Também é apresentado as questões de pesquisa que se buscam solucionar, bem como a execução dos experimentos e análise dos resultados.

- E por fim no Capítulo 6, **Considerações Finais e Trabalhos Futuros**, apresenta-se as considerações finais, análise das atividades já desenvolvidas, bem como o que é pretendido para trabalhos futuros.

2 Fundamentos Teóricos

Este capítulo apresentará conceitos relevantes da tecnologia que será aplicada e que está relacionada ao desenvolvimento deste trabalho, tais como: Aprendizado de máquina, *deep learning*, transformação de imagens, e uma visão ampla sobre os *frameworks* comumente utilizados para implementação de GANs.

2.1 Inteligência Artificial (IA)

A definição precisa sobre o conceito de IA possui muitas definições para o mesmo tema, porém com sentidos bem semelhantes. Por exemplo, em um trabalho realizado por (KOK et al., 2009), os autores nos apresentam quatro definições, sendo elas:

1. Uma área de estudo na área da ciência da computação. A inteligência artificial está preparada com o desenvolvimento de se envolver em processos de pensamento semelhantes aos humanos, como aprendizado, raciocínio e autocorreção.
2. O conceito de que as máquinas podem ser melhoradas para assumir algumas capacidades normalmente consideradas como inteligência humana, como aprendizagem, adaptação, autocorreção, entre outras.
3. A extensão da inteligência humana por meio do uso de computadores, como no passado o poder físico foi estendido por meio do uso de ferramentas mecânicas.
4. Em um sentido restrito, o estudo de técnicas para usar computadores de forma mais eficaz por meio de técnicas de programação aprimoradas.

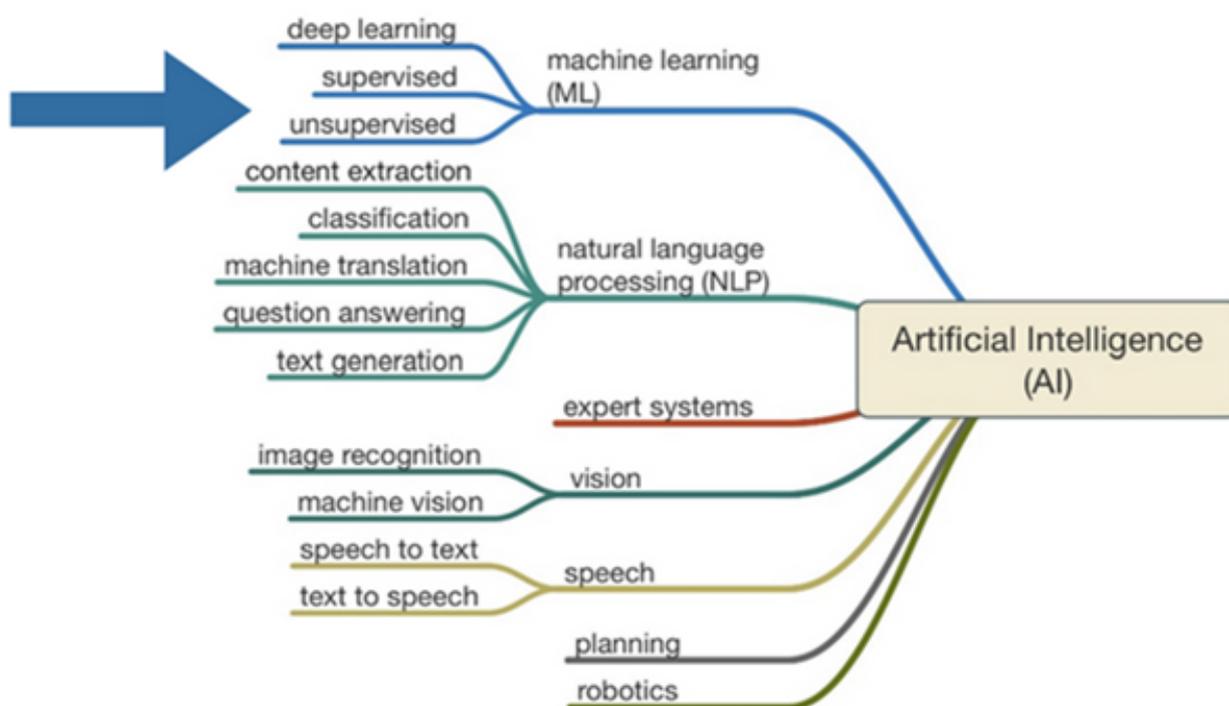
Para uma definição mais formal podemos citar os autores Luger (2004), Rich e Knight (1994) que argumentam que a IA é o ramo da ciência da computação que se ocupa com o comportamento inteligente. Ainda segundo Rich e

Knight (1994), IA é o estudo de como fazer os computadores realizarem coisas melhor que, atualmente, os humanos fazem.

A inteligência artificial também é um campo bastante voltado para o estudo e desenvolvimento de tecnologias em seus vários ramos ou linhas de pesquisa. A Figura 3, apresenta mais explicitamente essas linhas, sendo que dentre estes ramos, neste trabalho segue a linha de Aprendizado de Máquina (*Machine Learning*) com seu subcampo de Aprendizado Profundo (*Deep Learning*) e também a linha de visão computacional com seu subcampo de reconhecimento de imagens.

Durante as seguintes seções abordaremos sobre as linhas da IA que este projeto segue, sendo estas, *Machine Learning*, *Deep Learning* e visão computacional, e assim adentraremos no tema de uma das aplicações da IA, as GANs.

Figura 3 – Ramos da Inteligência Artificial.



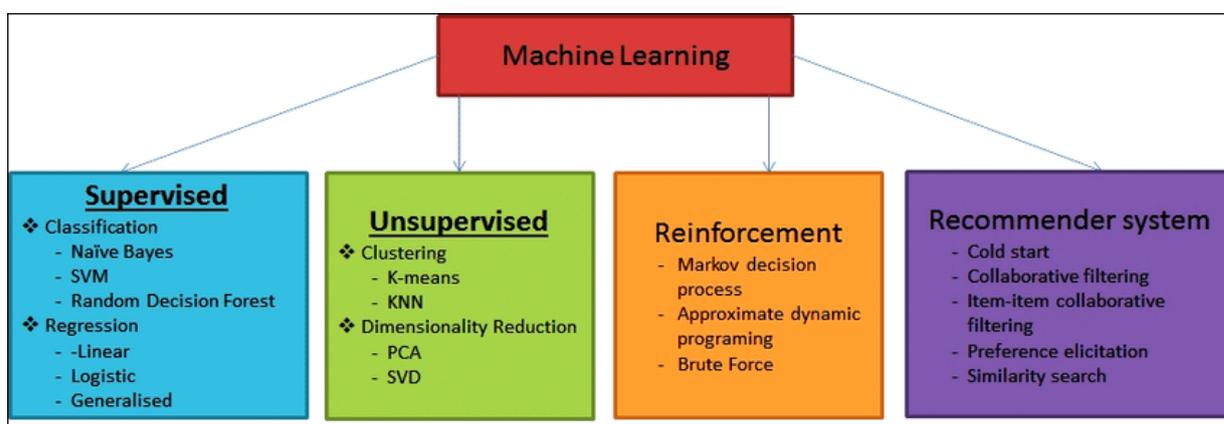
Fonte: (COELHO, 2020)

2.1.1 Aprendizado de Máquina

Uma definição para o este tema, possuindo 3.753 citações de acordo com o *Google Scholar*, foi dada por Arthur Samuel, em 1959, que foi um dos pioneiros no aprendizado de máquina. Ele então realiza a seguinte definição: campo de estudo que dá aos computadores a habilidade de aprender sem serem explicitamente programados (SAMUEL, 1959).

Neste campo de pesquisa, para obter resultados confiáveis explora-se a construção de algoritmos capazes de aprender com seus erros e fazer previsões sobre os dados a partir de algumas abordagens de aprendizagem, sendo as mais comumente usadas: supervisionada, semi-supervisionada, não supervisionada e reforço (AWAD; KHANNA, 2015). Na Figura 4, podemos visualizar alguns exemplos de algoritmos supervisionados, não supervisionados e de reforço.

Figura 4 – Exemplo de algoritmos para cada tipo de aprendizado.

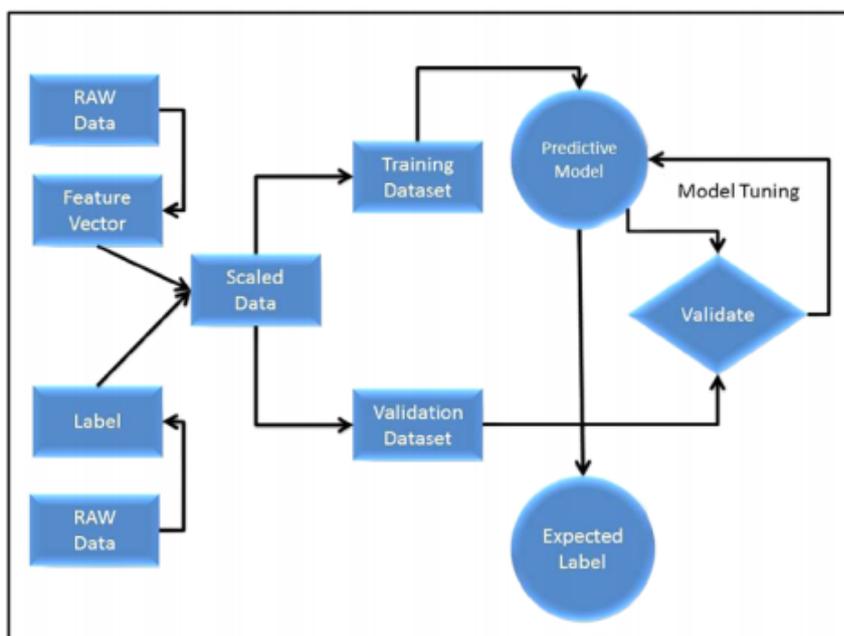


Fonte: (KUMAR et al., 2018)

Os algoritmos baseados em aprendizagem supervisionada usam os dados de treinamento rotulados para sintetizar a função do modelo que tenta generalizar a relação subjacente entre os vetores de entrada e os sinais de saída. Modelos bem treinados por estes algoritmos podem prever com precisão os rótulos de classe para fenômenos ocultos incorporados por instâncias de dados desconhecidos ou não observados (AWAD; KHANNA, 2015). Na Figura 5, pode ser observado o que foi descrito acima através de uma representação de

fluxo. Em que, a partir dos dados não tratados, modelos bem treinados e/ou aprimorados são capazes de prever rótulos esperados.

Figura 5 – Fluxo de aprendizagem supervisionada.



Fonte: (AWAD; KHANNA, 2015)

Algoritmos baseados em aprendizagem não supervisionada são projetados para descobrir estruturas ocultas em conjuntos de dados não rotulados, nos quais a saída desejada é desconhecida. Este tipo de aprendizagem envolve o treinamento por meio de modelos de dados probabilísticos. Além disso, dois dos exemplos mais populares para esta abordagem são o de agrupamento e redução de dimensionalidade.

Já a aprendizagem semi-supervisionada usa uma combinação de dados rotulados e não rotulados para gerar uma função de modelo ou um classificador. Esta metodologia opera entre as diretrizes da aprendizagem não supervisionada e da supervisionada (AWAD; KHANNA, 2015).

Por fim, métodos baseados em aprendizado de reforço envolvem a exploração de uma sequência adaptativa de ações ou comportamentos por um agente inteligente com a motivação de maximizar a recompensa cumulativa. Esta metodologia pode ser vista como um paradigma de aprendizado por tenta-

tiva e erro, com recompensas e punições associadas a uma sequência de ações. O agente de reforço possui algumas características como, mudar suas políticas com base em experiências coletivas e consequentes recompensas, e buscar ações passadas que explorou que resultam em recompensas (AWAD; KHANNA, 2015).

2.1.2 *Deep Learning*

O aprendizado profundo ou do inglês *deep learning*. É uma tecnologia da inteligência artificial, sendo uma ramificação do aprendizado de máquina. Ela se baseia no uso de tecnologias chamadas de redes neurais, com a proposta de melhorar o desempenho na análise de dados (CHOLLET, 2017).

Goodfellow et al. (2016), afirma que o desenvolvimento de *deep learning* foi motivado em parte pela falha dos algoritmos tradicionais em generalizar as tarefas de IA. O autor também afirma que o *deep learning* foi projetado para superar alguns obstáculos, como a generalização de dados ao se trabalhar com dados de alta dimensão e a insuficiência para aprender funções complicadas nesses espaços de altas dimensões, tendo em vista que estes espaços costumam consumir altos custos computacionais. Alguns exemplos de técnicas que visam resolver alguns destes obstáculos é o *data augmentation* e otimizações de regularizações como a *Smoothness Regularization*.

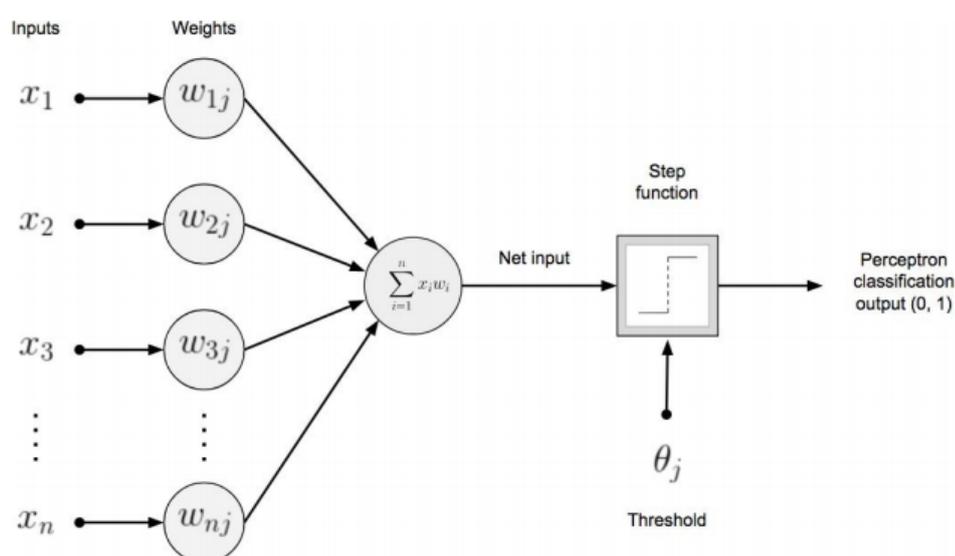
A maioria dos algoritmos de *deep learning* é baseada em um algoritmo de otimização chamado descida gradiente estocástica (do inglês Stochastic Gradient Descent - SGD). Essa técnica, SGD, é a principal forma de treinar grandes modelos lineares em conjuntos de dados muito grandes. A construção de algoritmos de *deep learning* acontece ao combinar uma especificação de um determinado conjunto de dados, uma função de custo, um procedimento de otimização e um modelo (GOODFELLOW et al., 2016).

A questão da descida do gradiente estocástico é que o gradiente é uma expectativa, cuja qual pode ser estimada aproximadamente usando um pe-

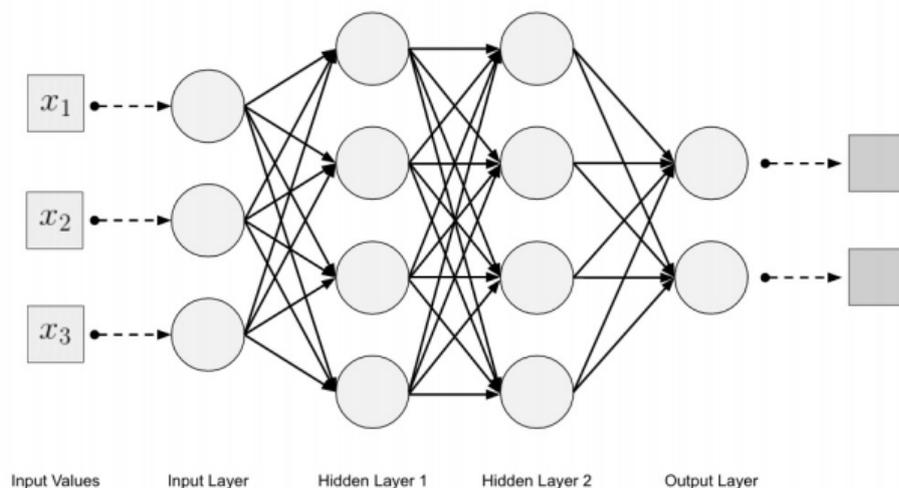
queno conjunto de amostras. De forma mais específica, em cada etapa do algoritmo, podemos amostrar um minibatch que fica responsável por pegar pequenos conjuntos de treinamento, onde seu tamanho é escolhido para ser executado com um número pequeno de amostras entre 1 a algumas centenas (GOODFELLOW et al., 2016). Esse tipo de técnica é um facilitador do *deep learning* ao se trabalhar com grandes conjuntos de dados.

Existem dois tipos de estruturas básicas para as redes neurais, as de camada única (*single-layer*) ou também conhecidas como rede perceptron e multicamadas (*multi-layer*), também conhecidas como rede *feed-forward*. Na rede de camada única, um conjunto de entradas é mapeado diretamente para uma saída usando a variação generalizada de uma função linear (AGGARWAL, 2018). Já as redes multicamadas, os neurônios são organizados em várias camadas, nas quais as camadas de entrada e saída são separadas por um grupo de camadas ocultas (AGGARWAL, 2018). Em *deep learning*, normalmente são mais utilizadas as redes *feed-forward* (PATTERSON; GIBSON, 2017). Podemos observar um exemplo de estrutura *single-layer* na Figura 6 e outro de *multi-layer* na Figura 7.

Figura 6 – Exemplo de estrutura *single-layer perceptron*.



Fonte: (PATTERSON; GIBSON, 2017)

Figura 7 – Exemplo de estrutura *multi-layer feed-forward* totalmente conectada.

Fonte: (PATTERSON; GIBSON, 2017)

Segundo Yamashita et al. (2018), o algoritmo mais estabelecido entre os vários modelos existentes de *deep learning*, é a rede neural convolucional, do inglês *convolutional neural network* (CNN). As redes CNN são um tipo de modelo oriundo do *deep learning* para processamento de dados sequenciais que possuem padrões de grades, como as imagens. Esse tipo de rede é uma construção matemática que normalmente é composta por três tipos de camadas, sendo essas: camada de convolução, *pooling* e camadas totalmente conectadas.

2.1.3 Generative Adversarial Networks (GANs)

As GANs são uma classe da inteligência artificial pertencente a uma das especializações de modelos generativos de *machine learning*. Essa tecnologia oferece algumas vantagens, como desempenhar um excelente papel ao gerar amostras de dados de forma eficiente e com boa compatibilidade a sua arquitetura que se baseia em redes neurais. As qualidades mencionadas, fizeram com que as GANs ganhassem destaque em alguns campos de atuação, como: Visão computacional, geração de imagens, tradução de imagens, transformação de imagens e super resolução (WANG et al., 2020).

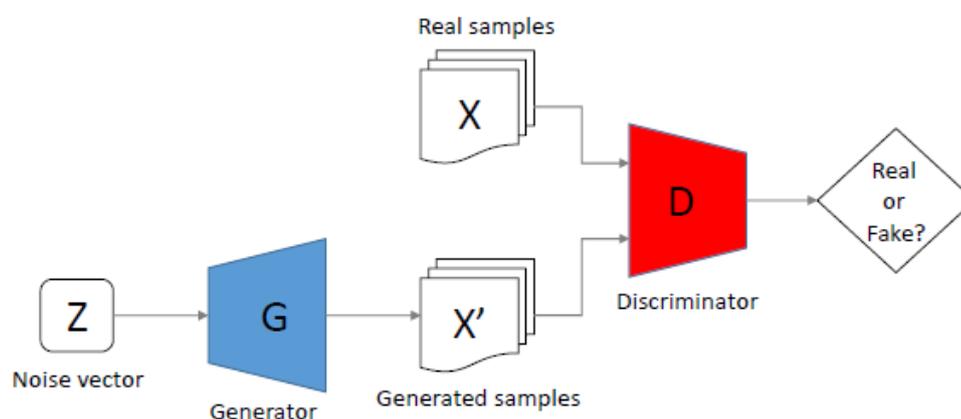
A arquitetura de uma GAN mais comumente conhecida é composta por duas redes neurais adversárias, o discriminador e o gerador. O funcionamento dessas GANs pode ser descrito da seguinte forma: Tomando como base as duas redes adversárias. O gerador é o responsável por gerar novas instâncias de dados, onde o objetivo é sempre gerar estes dados cada vez melhores. Já o discriminador é responsável por decidir se cada uma destas instâncias de dados que ele analisa pertence ou não ao conjunto de dados real de treinamento. Ele então obtém imagens reais e falsas do gerador e ao analisa-las, retorna um número entre 1 e 0, sendo o mais próximo de 1 representando seu reconhecimento de imagem autêntica e 0 representando o reconhecimento como falsas (GOODFELLOW et al., 2016). Em outras palavras, o discriminador e o gerador, são treinados simultaneamente no estágio de aprendizado. O discriminador então é otimizado para distinguir imagens reais de falsas que foram geradas pela rede geradora e o gerador é treinado para gerar imagens realísticas o suficiente para enganar o discriminador (WANG et al., 2020).

Ao longo dos anos decorridos de estudo das GANs diversas contribuições foram feitas por pesquisadores em uma vasta área de probabilidade de pesquisa, como geração de imagem, vídeo e voz. Além destas que são as mais tradicionalmente conhecidas, também temos contribuições na geração de personagens de animes (JIN et al., 2017), objetos 3D (WU et al., 2017a), trocas de rostos (DING et al., 2019), envelhecimento facial, criação de vídeos fake (NGUYEN et al., 2020), quadros artísticos (YI et al., 2019), dar vida a criações artísticas (ZAKHAROV et al., 2019) e além de muitas outras aplicações.

Outras definições de GAN podem ser observadas em algumas pesquisas existentes no estado da arte, como no trabalho de Wu et al. (2017b), os autores buscam como foco desta pesquisa, apresentar a definição de uma GAN, mostrar outros trabalhos correlatos discutindo ideias, contribuições e aplicações. No trabalho, também foi apresentado aplicativos para auxiliar na síntese de imagem e ainda realizaram uma breve discussão sobre os desafios e limitações para os métodos que são baseados em GANs.

Na Figura 8 que foi apresentada no trabalho de Wu et al. (2017b) é definido o que são e qual é o funcionamento destas GANs. Desta forma, uma GAN contém duas redes, sendo elas um gerador G e um discriminador D . Este gerador tenta criar imagens falsas, mas plausíveis, enquanto o discriminador tenta distinguir se as imagens geradas pelo gerador são falsas. Em outras palavras o gerador G mapeia um vetor de ruído Z para uma imagem, sendo $G(Z) \rightarrow X$, e o discriminador D é definido como $D(x) \rightarrow [0, 1]$, que fica encarregado de classificar como uma imagem real ou falsa, em ambos os casos o resultado deve ser perto de 1.

Figura 8 – Estrutura de GANs.



Fonte: (WU et al., 2017b)

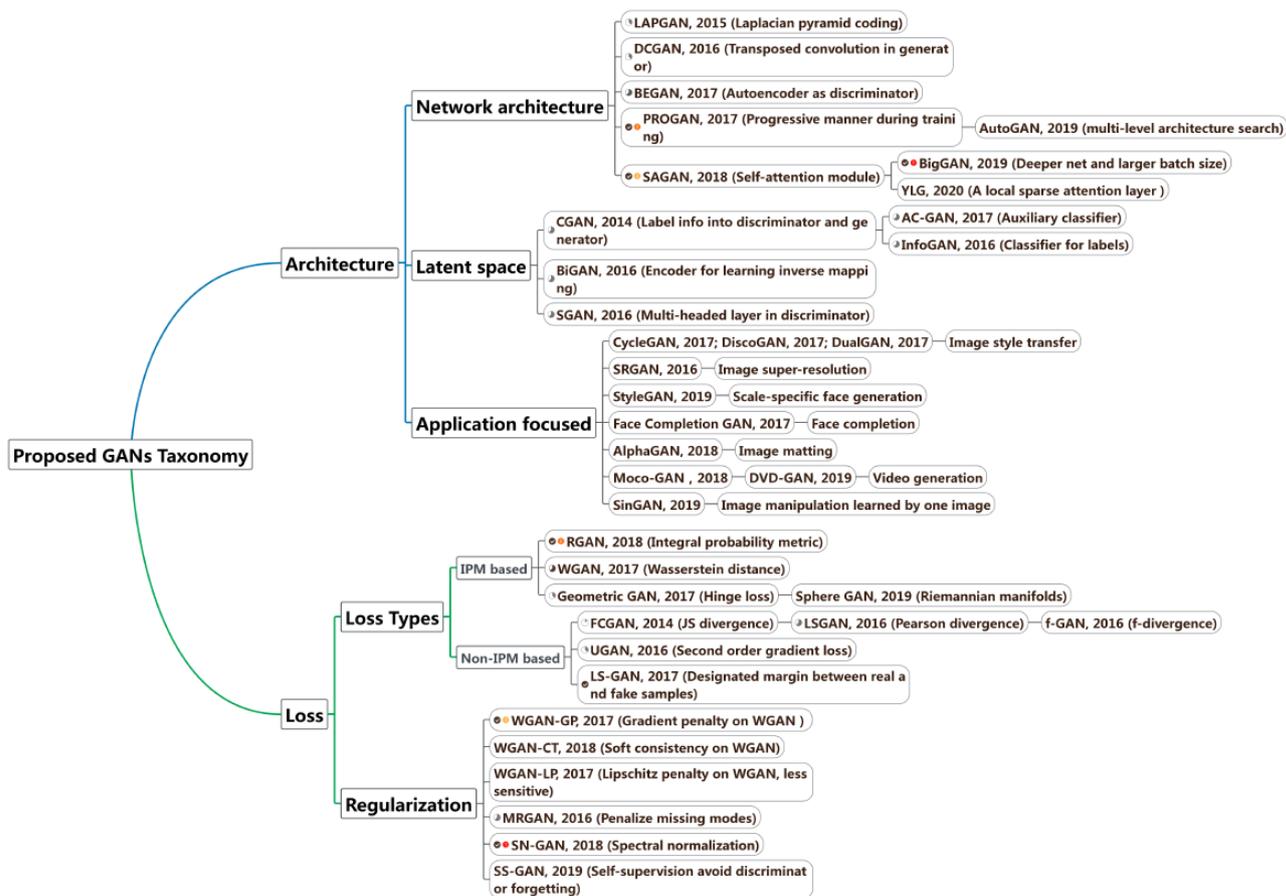
2.1.3.1 Taxonomia de modelos generativos

As taxonomias de modelos generativos nada mais são do que variantes de GANs subdivididas em uma determinada categoria, tais como: cGAN que está subdividida pela categoria de espaço latente; DCGAN, que está na categoria de arquitetura de rede; e SRGAN, que está subdividida na categoria de foco da aplicação. Esta subdivisão pode ser melhor observada pela Figura 9. No trabalho realizado por (WANG et al., 2020), os autores propõem uma taxonomia na qual eles subdividem as GANs atuais em duas variantes, a da arquitetura e

a de perda.

Na primeira, de arquitetura, é subdividida em três categorias, de rede, espaço latente e foco da aplicação. A categoria de rede refere-se a alguma melhoria ou modificação realizada na arquitetura em questão. Espaço latente refere-se a modificações em diferentes espaços latentes. E a categoria focada na aplicação refere-se à modificação feita de acordo com diferentes aplicações. Já as variantes de perda, é subdividida em tipos de perda e regulação. Os tipos de perda referem-se a diferentes funções de perda (tais como, métricas de probabilidade integral (IPM) e não baseada em IPM) a serem otimizadas para GANs, e a regulação refere-se à penalização adicional projetada para a função de perda ou qualquer tipo de operação de normalização feita na rede. Essa taxonomia apresentada pelos autores pode ser observada de acordo com a Figura 9.

Figura 9 – Taxonomia proposta pelos autores.



Fonte: (WANG et al., 2020)

2.1.3.2 AutoEncoders

Os autoencoders são redes neurais não supervisionadas, esse tipo de rede é composto por duas partes, codificador (*Encoder*) e decodificador (*Decoder*). Os autoencoders são treinados para preservar o máximo de informações possíveis. Ou seja, eles aprendem a compactar e codificar os dados de maneira eficiente, feito isso, aprendem a como reconstruir esses dados de volta a partir de uma representação codificada para uma que seja o mais próximo possível da entrada original. (GOODFELLOW et al., 2016).

Dito isso, as partes do autoencoder podem ser descritas da seguinte forma. O *encoder* é a parte responsável por compactar a entrada de um dado em

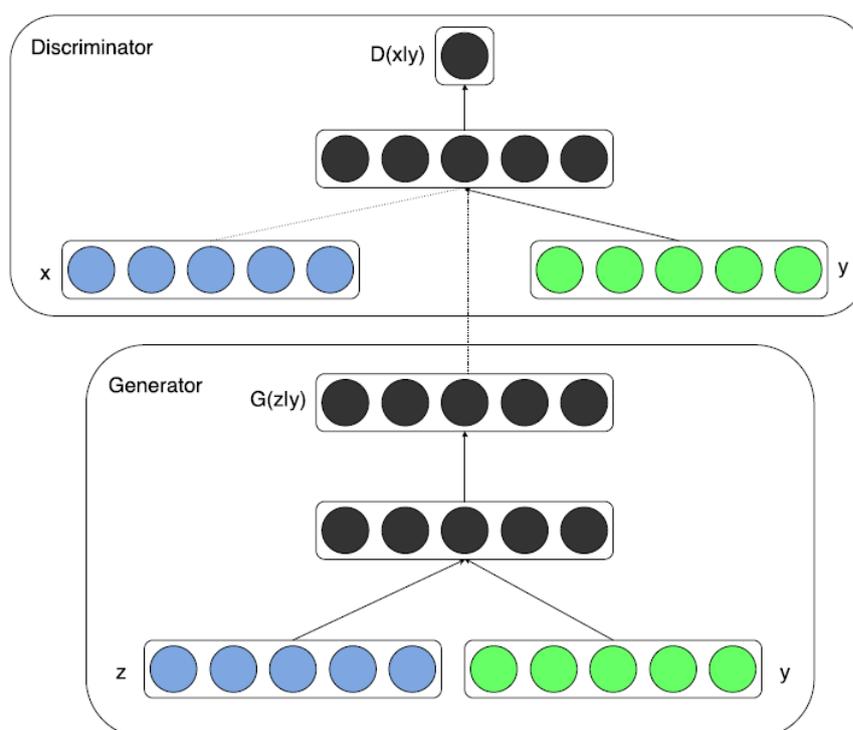
uma representação do espaço latente. A função de codificação que o representa é dada por $h = f(x)$, onde x seria o dado original de entrada e h é a representação latente. Já o *decoder* é responsável por reconstruir a entrada da representação do espaço latente. A função do decodificador é dada por $r = g(h)$, onde r é o dado de saída reconstruído, e h é a representação do estado latente codificado, ou seja o dado inicial de entrada codificado (GOODFELLOW et al., 2016).

2.1.4 GANs Condicionais

As GANs condicionais, são uma extensão dos modelos tradicionais, isso se tanto o gerador quanto o discriminador estiverem condicionados a uma informação y extra qualquer de entrada. O tipo desta informação auxiliar adicionada pode ser rótulos de classe ou dados de outras modalidades (MIRZA; OSINDERO, 2014).

A informação y normalmente é codificada dentro do gerador e discriminador antes de ser concatenado com as outras informações z , vetor de ruído de entrada, codificado para o gerador e x codificado para o discriminador. Isto pode ser melhor observado na Figura 10, onde apresenta uma ilustração de uma estrutura de GAN condicional simples (WANG et al., 2020).

Figura 10 – Ilustração de uma estrutura simples de GAN condicional.



Fonte: (MIRZA; OSINDERO, 2014)

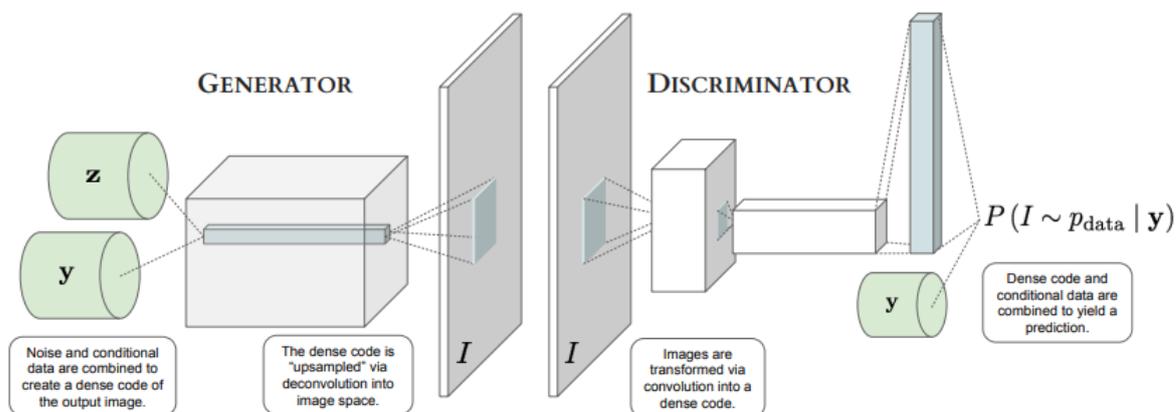
Para que se entenda melhor a arquitetura e o funcionamento de uma cGAN, apresentaremos sua estrutura comumente usada expandida. Utilizaremos a Figura 11 apresentada por (GAUTHIER, 2014), para explicar a estrutura.

As *conditional generative adversarial networks (cGANs)*, possuem duas estruturas básicas que compõem a rede, estas são o gerador e o discriminador. Segundo Gauthier (2014), o gerador é uma rede neural deconvolucional e o discriminador uma rede convolucional.

O funcionamento dessa estrutura tradicional consiste em primeiro o vetor de ruído z e a informação condicional y são combinados para criar um denso código da imagem de saída, em seguida o código gerado é filtrado e expandido através da deconvolução no espaço da imagem, as imagens então são transformadas no discriminador através da convolução em um denso código, que é combinado a dados condicionais para assim produzir a previsão

(GAUTHIER, 2014). A estrutura expandida pode ser observada pela Figura 11.

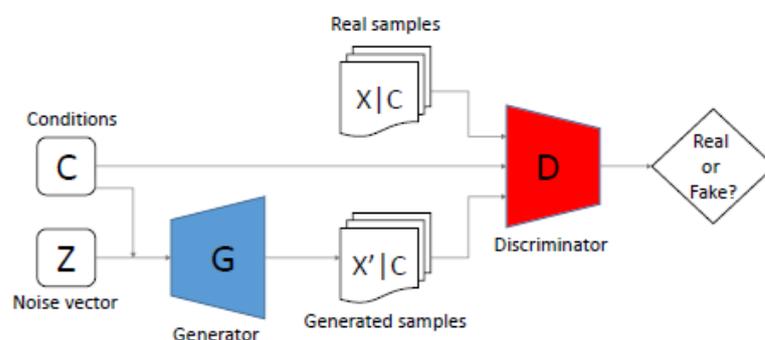
Figura 11 – Arquitetura da estrutura expandida de uma cGAN.



Fonte: (GAUTHIER, 2014)

No trabalho de Wu et al. (2017b), existente no estado da arte, os autores apresentam alguns tipos de propostas e derivações de GANs existentes. E uma destas GANs propostas foi a cGAN, sua estrutura é apresentada na Figura 12. Nela, o gerador G e o discriminador D são condicionados a alguma informação extra c , essas podem ser rótulos de classe, texto ou esboço. Desta forma, c é passado como informação adicional para G e D. Esse tipo de estrutura fornece controle adicional sobre os tipos de dados que estão sendo gerados, diferente do GAN convencional que não tem esse tipo de controle. Isso faz com que as cGANs sejam populares para aplicações de síntese e edição de imagens.

Figura 12 – Estrutura de cGANs.



Fonte: (WU et al., 2017b)

Um exemplo prático de uso das estruturas de cGAN, é o trabalho realizado por Antipov et al. (2017), que possui como objetivo a geração de faces envelhecidas. A estrutura baseada em cGAN proposta pelo autor, consiste em primeiro realizar a entrada de um determinado dado de entrada (imagem) x com uma idade y_0 , estas informações então serão passadas a um *encoder E*, que ficará responsável por encontrar um ótimo vetor latente z^* que permitirá a reconstrução de uma imagem sintética com idade y_0 . A última etapa do processo consiste na passagem de uma idade alvo para predição e do vetor latente z^* , obtido anteriormente, para o gerador G que ficará responsável juntar os dados coletados e produzir a previsão (ANTIPOV et al., 2017). O trecho descrito pode ser observado pela Figura 13.

2.2 Frameworks para implementação de GANs

Os *frameworks* existentes mais comuns para a implementação de GANs, que é o nosso foco, são: Tensorflow, Keras, PyTorch e Caffe (KAPOOR et al., 2019; VASILEV et al., 2019; Cengil et al., 2017; JIA et al., 2014). O tensorflow é um *framework* de código aberto desenvolvido pela equipe do *Google Brain* para uso em *deep learning*. A ferramenta suporta várias das linguagens populares como Python, C++, Java, R e Go. Além disso conta com a API Keras, integrada ao

Tensorflow, que especifica como os componentes de software devem interagir (KAPOOR et al., 2019).

Keras é uma API de código aberto escrita em Python para criar e treinar modelos de deep learning, foi desenvolvida com foco em permitir a experimentação rápida com redes neurais profundas (KAPOOR et al., 2019; KERAS, 2021). Esta API pode ser integrada a vários tipos de *frameworks* voltados para o campo de *deep learning* como Amazon MxNet, Theano, Microsoft CNTK e o próprio Tensorflow como comentado anteriormente (KAPOOR et al., 2019). Além de tudo, a ferramenta permite que o mesmo código seja executado na CPU ou GPU. Possui suporte para as redes convolucionais (CNN), as redes recorrentes (RNN) e qualquer combinação que seja com as mesmas. Também suporta arquiteturas de redes arbitrárias, ou seja, modelos de múltiplas entradas ou saídas, compartilhamento de camadas e modelos, entre outros (KERAS, 2021).

PyTorch é um *framework* em Python de código aberto voltado para *deep learning*, foi baseado no Torch e desenvolvida principalmente pelo Facebook. Este software nos oferece alguns recursos como uma GPU, operação de matriz multidimensional acelerada ou tensor, e gráficos computacionais, para que se possa construir redes neurais (VASILEV et al., 2019).

O Caffe é um *framework* de código aberto voltado ao *deep learning*, ele foi desenvolvido e é mantido pelo Berkeley Vision and Learning Center (BVLC) (Cengil et al., 2017; JIA et al., 2014). O software pode ser utilizado para desenvolver aplicações do mundo real (Cengil et al., 2017). Seu código é escrito na linguagem C++, com CUDA usado para computação de GPU e possui ligações quase completas, mas bem suportadas para a linguagem Python e Matlab (JIA et al., 2014). Além disso, fornece testes de unidade para precisão, velocidade para instalação, e busca seguir as boas práticas de engenharia de software. Outro fator que chama bastante a atenção para este software é a capacidade de modularidade cuidadosa do código e a separação clara da definição de rede da implementação real (Cengil et al., 2017; JIA et al., 2014).

2.3 Destaques conceituais para pesquisa

Durante este capítulo apresentamos conceitos importantes para entendermos um pouco em que área se encaixa esta pesquisa. Nele entendemos a relação que existe entre as GANs, *deep learning*, *machine learning* e inteligência artificial.

Falando a respeito de arquitetura, as GANs normalmente utilizando autoencoders e a seção a qual explicitamos os conceitos sobre o assunto, auxiliam na compreensão do mesmo. Como vimos, esses autoencoders são divididos em duas partes, sendo elas o codificador (*encoder*) e o decodificador (*decoder*), é aí que entra a forte relação com a pesquisa proposta, pois serão dois componentes essenciais para a arquitetura da GAN.

Apresentamos também conceitos sobre *data augmentation* que é uma técnica muito valiosa em relação a dificuldade que se tem de trabalhar com a quantidade e diversidade de dados escassos. Com isso, a técnica vem a socorro para auxiliar na geração de dados artificiais para aumentar essa quantidade e diversidade. Tomando em consideração os benefícios do *data augmentation*, essa é uma técnica que com certeza, se necessário, virá em auxílio durante este trabalho.

Por fim, também apresentamos alguns exemplos de frameworks voltados para a implementação de GANs. Estes frameworks são estruturas necessárias para a implementação de um projeto de pesquisa, que irão de auxílio na implementação da arquitetura das GANs. Com isso podemos ter uma breve noção sobre os tipos de *framework* que podemos usar, como o Tensorflow e Keras.

3 Trabalhos Correlatos

Neste capítulo são apresentados alguns trabalhos de pesquisa que foram desenvolvidos usando técnicas com GAN. Neles, são mostrados exemplos de aplicações de GANs e casos semelhantes.

3.1 Trabalho 1 - *Face aging with conditional generative adversarial networks*

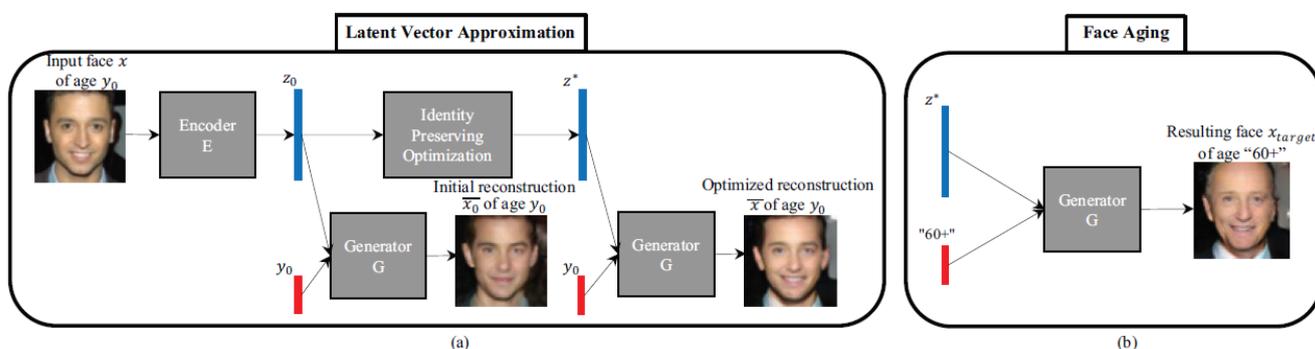
No trabalho proposto por Antipov et al. (2017) é proposto um método baseado em envelhecimento facial, um tipo de rede adversária generativa, do inglês, *adversarial generative network (GAN)*, onde o principal objetivo é gerar ou retroceder um rosto em uma idade futura ou passada preservando o máximo possível de características.

A proposta de Antipov et al. (2017) chama-se Age-cGAN, que nada mais é que uma rede geradora condicional generativa de idade. Segundo os autores esse tipo de derivação de GAN desenvolvido por eles, foi o primeiro a gerar imagens sintéticas de alta qualidade de acordo com as idades. Além disso, propuseram uma abordagem de uma otimização de vetores latentes, o que possibilita que o *framework* desenvolvido reconstrua as imagens de entrada em imagens sintéticas, com as características preservadas da pessoa.

No modelo proposto por Antipov et al. (2017), o envelhecimento é feito em duas etapas. Onde no primeiro, como ilustrado na Figura 13 - (a), deve-se dar uma entrada com uma imagem x de uma idade y_0 , esta imagem será passada para o *encoder* E , com a função de encontrar um ótimo vetor latente z^* , que permitirá a reconstrução de uma imagem sintética \bar{x} de uma idade y_0 o mais próximo o possível da imagem de entrada. Já na segunda etapa, como mostrado na Figura 13 - (b) será passado de entrada a idade alvo, y_{target} , juntamente ao vetor latente para o Gerador G , com isso irá gerar continuamente o rosto

resultante simplesmente ao passar a variável y_{target} o valor da idade desejada para a imagem sintética.

Figura 13 – Arquitetura do Age-cGAN proposto pela equipe.



Fonte: (ANTIPOV et al., 2017)

O dataset utilizado para treinamento do Age-cGAN de Antipov et al. (2017), foi o IMDB-Wiki cleaned (ANTIPOV et al., 2016). De acordo com os autores, este dataset possui 120 mil imagens que é um subconjunto do dataset público IMDB-Wiki (ROTHER et al., 2015). Mais precisamente, 110 mil imagens foram utilizadas para o treinamento do Age-cGAN e os demais 10 mil foram utilizados para avaliação da preservação de identidade da reconstrução facial.

Para obter os resultados, Antipov et al. (2017) utilizaram o software OpenFace (AMOS et al., 2016) que é uma aplicação de reconhecimento facial de código aberto. Onde, a partir de duas imagens, ele fica responsável por decidir se aquelas imagens são a mesma pessoa ou não. Com isso, eles aplicaram diferentes tipos de reconstrução propostas pelos autores: *Initial Reconstruction* (Z_0), *Pixelwise Optimization* (Z_{pixel}^*) e *Identity-Preserving Optimization* (Z_{IP}^*).

- *Initial Reconstruction* (Z_0), que é uma reconstrução inicial de imagens geradas usando aproximações iniciais latentes, neste a pontuação obtido foi de 53.2%.
- No segundo, *Pixelwise Optimization* (Z_{pixel}^*), imagens geradas reconstruídas usando as opções latentes otimizadas, o *Pixelwise* e *Identity-Preserving*. Uma ligeira melhoria foi notada com score de 59.8%.

- Já no terceiro e último, *Identity-Preserving Optimization* (Z_{IP}^*), envelhecimento de imagens geradas que foram reconstruídas utilizando o *Identity-Preserving* Z_{IP}^* , aproximações latentes e condicionadas às respectivas categorias de faixa etária. Com esta otimização final e proposta por este trabalho, obteve-se um excelente resultado com score de 82.9%.

3.2 Trabalho 2 - *Attribute-aware Face Aging with Wavelet-based Generative Adversarial Networks*

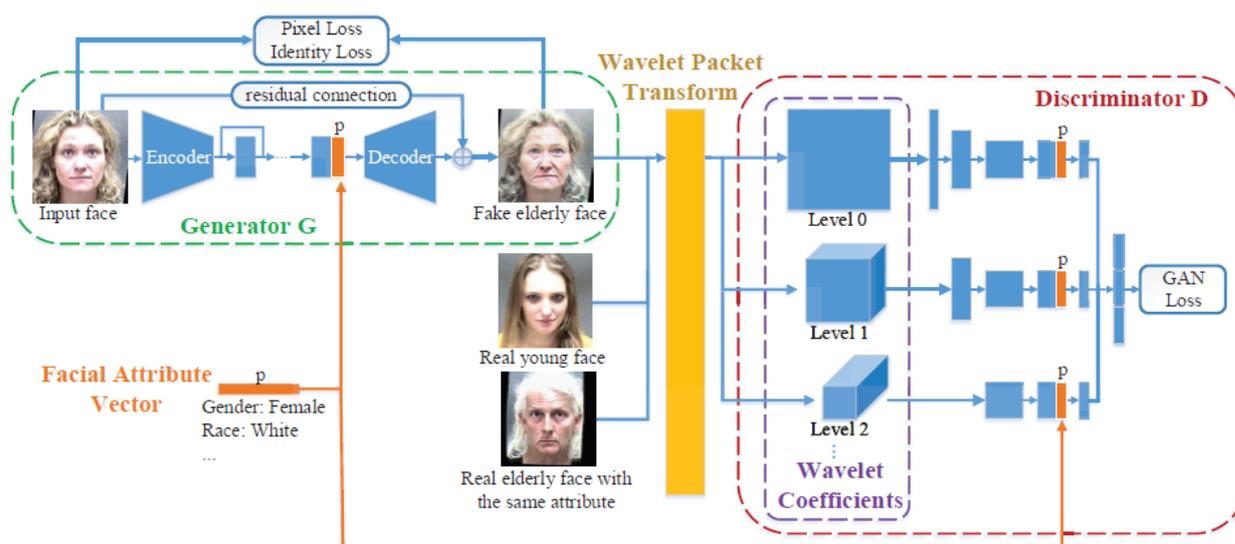
Segundo Liu et al. (2019), existem trabalhos que recorrem a conjuntos de dados não emparelhados para aprender o mapeamento de idades, porém podem existir consequências do uso destes dados. Tendo em mente, que podem gerar ambiguidades de correspondência de imagens, e alterações não naturais de atributos faciais durante o processo de envelhecimento.

Liu et al. (2019) propuseram um modelo de envelhecimento facial com o uso de *Generative Adversarial Networks (GANs)* baseadas em *wavelets*. No modelo foi incorporado vetores de atributos faciais, no gerador e no discriminador, para que cada imagem de rosto sintetizada seja o mais fiel possível ao atributo de entrada correspondente, no caso a face de origem. Um módulo de processo de decomposição de informação (do inglês *wavelet packet transform - WPT*), também foi incorporado com finalidade de extrair detalhes da imagem em várias escalas para gerar detalhes com efeitos realísticos de envelhecimento.

A arquitetura do *framework* desenvolvido por Liu et al. (2019) é descrita de forma objetiva na imagem abaixo, Figura 14. A estrutura é composta por um gerador G, um discriminador D, um módulo de decomposição de informação e por um vetor de atributo facial. No gerador G, ele aprende o mapeamento da idade e gera imagens de face envelhecidas realísticas. O discriminador D, possui como função distinguir imagens de rostos sintetizadas genéricas, com base nos coeficientes *wavelet*, que são calculados pelo módulo de decomposição de informação. E por fim, o vetor de atributo p, que possui como função

descrever a imagem da face de entrada, e é incorporado ao gerador e ao discriminador com finalidade de reduzir ambiguidades aos dados de treinamento não emparelhados.

Figura 14 – Arquitetura do modelo de *GAN-Wavelet* proposto pela equipe.



Fonte: (LIU et al., 2019)

Para a aplicação e teste da arquitetura foram utilizados os bancos de dados MORPH (RICANEK; TESAFAYE, 2006) e CACD (CHEN et al., 2015). No primeiro, possui uma base de dados de 55 mil imagens de rostos de mais de 13 mil indivíduos, de acordo com (LIU et al., 2019), esta base é composta por imagens coloridas, expressões faciais neutras e uma boa iluminação com fundo simples, todas essas características torna essa base de dados menos complexa a ser analisada.

Em relação ao dataset CACD, de acordo com Liu et al. (2019), ele possui uma base de dados de 163.446 imagens de rostos de 2 mil celebridades capturadas em condições muito controladas, com variações de pose, iluminação e expressão, o que torna essa base de dados mais desafiadora de ser analisada.

Sobre os detalhes de implementação, o grupo de pesquisadores usaram uma técnica chamada do inglês *Multi-task Convolutional Network (MTCNN)*, ou seja, rede convolucional multitarefa. Esta técnica foi usada para recortar e

alinhar as imagens de acordo com os pontos faciais detectados pela MTCNN, estas imagens foram divididas em grupos de faixa etária. Para avaliar o desempenho, Liu et al. (2019) utilizaram todas as métricas por meio de APIs públicas do Face++ (MEGVII-INC., 2021), e além disso escolheram o algoritmo Adam para ser o otimizador do Gerador do Discriminador.

Em relação aos resultados, os autores realizam uma comparação com o primeiro trabalho que foi baseado no Morph. Com essa análise puderam perceber que tratando dos métodos baseados em GAN, duas dessas se destacaram com uma melhoria significativa de qualidade na geração dos resultados, sendo essas a GLCA-GAN (LI et al., 2018) e a PAG-GAN (YANG et al., 2018) como mostrado na Tabela 1.

Tabela 1 – Resultados de estimativa de idade encontrados pela equipe em Morph e CACD (diferenças de idades médias são medidas em valor absoluto)

	Morph				CACD		
Grupo de idade	31 - 40	41 - 50	51 +	Grupo de idade	31 - 40	41 - 50	51 +
Distribuições de idades estimadas				Distribuições de idades estimadas			
Genérico	38.60	47.74	57.25	Genérico	38.51	46.54	53.39
Sintético	38.47	47.55	56.57	Sintético	38.88	47.42	54.05
Diferença da média de idades				Diferença da média de idades			
CAAE	10.08	15.49	21.42	CAAE	5.76	11.53	17.93
GLCA-GAN	0.23	3.61	8.61	GLCA-GAN	1.72	2.07	2.85
PAG-GAN	0.38	0.52	1.48	PAG-GAN	0.70	0.22	0.57
Ours	0.13	0.19	0.68	Ours	0.37	0.58	0.66

Fonte: (LIU et al., 2019)

Em outras duas verificações, Tabela 2 e Tabela 3, o método proposto se mostrou com um grande desempenho, podemos perceber claramente que o projeto em questão superou todos os demais utilizando o dataset Morph, já no CACD sua precisão ficou bem semelhante ao da PAG-GAN. Já na Tabela 3, o método foi muito eficiente, nos dois datasets propostos, superando as outras GANs comparadas.

Tabela 2 – Resultados encontrados pela equipe na verificação facial em Morph e CACD

Morph				CACD			
Grupo de idade	31 - 40	41 - 50	51 +	Grupo de idade	31 - 40	41 - 50	51 +
Confiança na verificação				Confiança na verificação			
30 -	95.77	94.64	87.53	30 -	93.67	91.54	90.32
31 - 40	-	95.47	89.53	31 - 40	-	91.74	90.54
41 - 50	-	-	90.50	41 - 50	-	-	91.12
Taxa de verificação (%)				Taxa de verificação (%)			
CAAE	15.07	12.02	8.22	CAAE	4.66	3.41	2.40
GLCA-GAN	97.66	96.67	91.85	GLCA-GAN	97.72	94.18	92.29
PAG-GAN	100.00	98.91	93.09	PAG-GAN	99.99	99.81	98.28
Ours	100.00	100.00	98.26	Ours	99.76	98.74	98.44

Fonte: (LIU et al., 2019)

Tabela 3 – Resultados encontrados pela equipe nas taxas de preservação de atributos faciais para 'Genero' e 'Raça' em Morph e CACD

Grupo de idade	Taxa de preservação (%) do 'Genero'						Taxa de preservação (%) da 'Raça'		
	Morph			CACD			Morph		
	31 - 40	41 - 50	51 +	31 - 40	41 - 50	51 +	31 - 40	41 - 50	51 +
GLCA-GAN	96.30	95.43	95.77	87.27	86.79	85.89	91.79	89.52	89.34
PAG-GAN	95.96	93.77	92.47	83.97	81.28	70.05	95.83	88.51	87.98
Ours	97.37	97.21	96.07	90.71	87.63	87.19	95.86	94.10	93.22

Fonte: (LIU et al., 2019)

3.3 Trabalho 3 - *Interpreting the Latent Space of GANs for Semantic Face Editing*

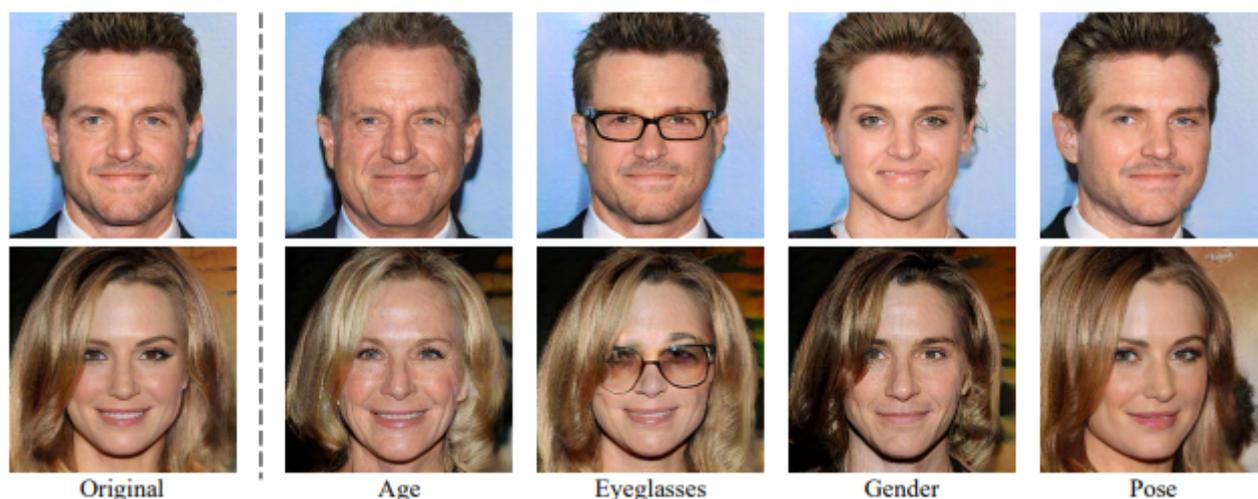
Shen et al. (2020) apresentam um framework de GANs para a interpretação de faces chamado de *Interpreting Face GANs* (InterFaceGAN). Esta aplicação possui como finalidade identificar a semântica codificada no espaço latente de modelos de síntese facial pré treinados e em seguida utilizá-los para a edição semântica da face.

Shen et al. (2020) conduziram estudos detalhados sobre como diferentes semânticas são codificadas no espaço latente de GANs para a síntese facial. E por meio destes estudos puderam constatar que o código latente de modelos generativos treinados, aprendem uma representação desembaraçada após transformações lineares. Shen et al. (2020) abordaram o desemaranhamento

entre diferentes semânticas, e perceberam que podemos desacoplar alguns atributos emaranhados, como os idosos serem mais propensos a usar óculos que os jovens, e isso foi percebido por meio de projeção do subespaço linear (SHEN et al., 2020).

Algumas das contribuições dadas Shen et al. (2020) foram, primeiramente a própria proposta da InterFaceGAN para explorar como a semântica única ou múltipla é codificada no espaço latente das GANs, observar que as GANs aprendem espontaneamente vários subespaços latentes correspondentes a atributos específicos. Outra contribuição, foi mostrar que a aplicação permite a edição semântica de faces, como a edição de gênero, idade, expressão, presença de óculos, pose do rosto com qualquer modelo de GAN pré-treinado.

Figura 15 – Manipulações de vários atributos faciais por meio da variação dos códigos latentes de um modelo de GAN bem treinado.



Fonte: (SHEN et al., 2020)

O dataset utilizado para treinamento dos modelos e estudo na InterFaceGAN de Shen et al. (2020), foi o CelebA (LIU et al., 2015) junto a rede ResNet50 (HE et al., 2016), que é um dataset público, disponível apenas para fins de pesquisa, com mais de 200 mil imagens faciais. Além deste, também foi usado o dataset CelebA-HQ, que é uma versão de alta qualidade do CelebA, criada e apresentada por Karras et al. (2018), a base de dados consiste em 30 mil

imagens com resolução de 1024x1024.

Durante a fase de avaliação da aplicação InterFaceGAN, Shen et al. (2020) utilizaram duas GANs em específico no estado da arte, sendo a PGGAN (KARRAS et al., 2018) e StyleGAN (KARRAS et al., 2019a). A PGGAN foi usada especificamente para interpretar o espaço latente do gerador tradicional, desta forma avaliando a separação e manipulação do espaço latente, bem como a manipulação condicional. Já a StyleGAN foi usada para investigar o gerador baseado em estilo e também comparar as diferenças entre os dois conjuntos de representações latentes no StyleGAN.

Com relação aos resultados, Shen et al. (2020) treinaram cinco SVMs lineares independentes em pose, sorriso, idade, gênero e óculos, em seguida, avaliaram 6 mil amostras no conjunto de validação, bem como em todo o conjunto de 480 mil amostras aleatórias. Com isso os pesquisadores descobriram em seus resultados, que todos os limites lineares atingem mais de 95% de precisão no conjunto de validação e mais de 75% em todo o conjunto. Os demais resultados de acurácia de classificação nos limites de separação do espaço latente podem ser melhor observados pela Tabela 4.

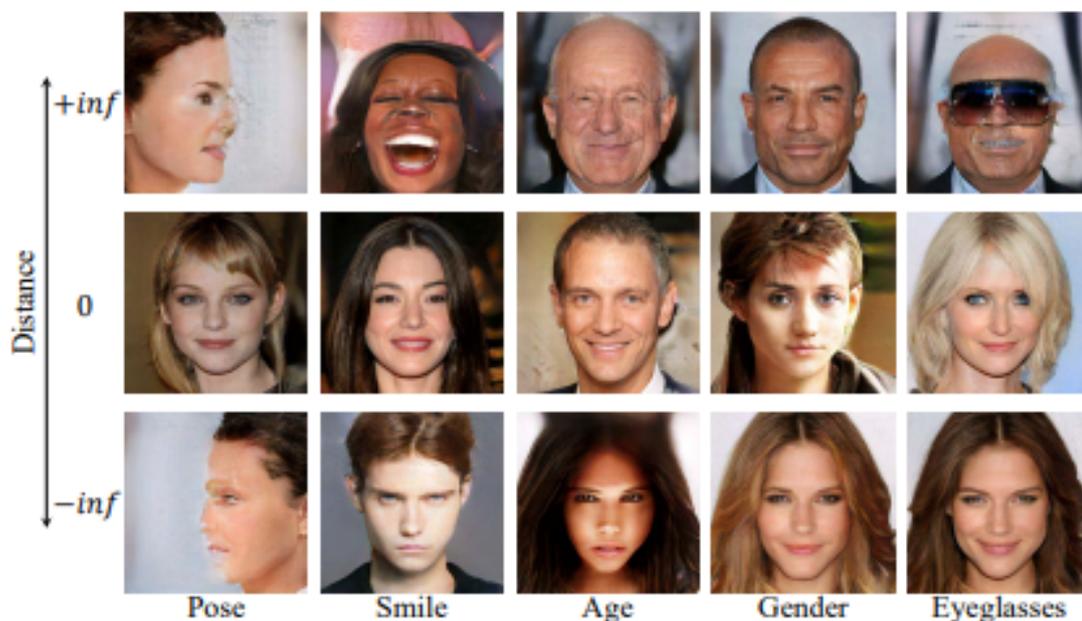
Tabela 4 – Resultados em porcentagem de acurácia de classificação nos limites de separação no espaço latente em relação a diferentes atributos.

Dataset	Pose	Sorriso	Idade	Gênero	Óculos
Validação	100.0	96.9	97.9	98.7	95.6
Todos	90.3	78.5	75.3	84.2	80.1

Fonte: (SHEN et al., 2020)

Observou-se também que ao manipular o código latente, através da separação do espaço latente, produz um efeito de distância em que as amostras sofrerão alterações severas na aparência se forem movidas para muito longe do limite de separação/decisão, e tenderão a se tornar casos extremos como na Figura 16. Já com a manipulação perto do limite de separação, pode-se obter resultados melhores.

Figura 16 – Amostras de síntese com a distância próxima (linha do meio) e extremamente longe (linha superior e inferior) do limite de separação, por atributo.

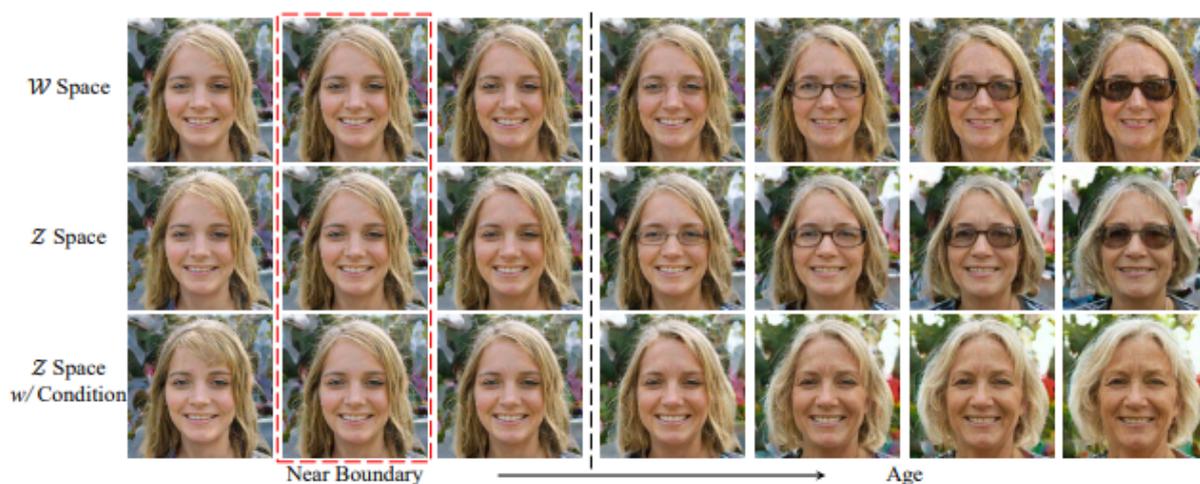


Fonte: (SHEN et al., 2020)

Ainda sobre os resultados, com base nos testes executados na StyleGAN, os pesquisadores descobriram que o espaço latente W de fato aprende uma representação mais desemaranhada, como apontado por Karras et al. (2019a). E esse desemaranhamento ajuda o espaço W a atingir uma forte superioridade sobre o espaço Z para a edição de atributos.

A comparação entre os resultados de edição de atributos dos espaços latentes podem ser observados pela Figura 17, onde, comparado ao espaço Z (segunda linha da figura), o espaço W (primeira linha da figura) tem melhor desempenho, principalmente na manipulação de longa distância. No entanto, podemos utilizar a manipulação condicional para descorrelacionar os dois atributos (idade e óculos) no espaço Z (terceira linha da figura), fazendo assim resultar em resultados mais atraentes. Porém, esta manipulação condicional, não pode ser aplicada ao espaço W (SHEN et al., 2020).

Figura 17 – Análise do espaço latente Z e do espaço latente desemaranhado W do StyleGAN tomando a manipulação da idade como exemplo.



Fonte: (SHEN et al., 2020)

3.4 Trabalho 4 - *A Style-Based Generator Architecture for Generative Adversarial Networks*

Karras et al. (2019a) apresentam uma arquitetura de gerador alternativa para redes adversárias geradoras, se baseando na literatura de transferência de estilo. Desta forma, os pesquisadores redesenharam a arquitetura do gerador de forma que possua novas maneiras de controlar o processo de síntese de imagem. E uma das formas de controlar é o gerador começando com uma entrada constante aprendida e ajusta o estilo da imagem em cada camada de convolução com base no código latente, desta forma controlando diretamente a força dos recursos da imagem em diferentes escalas (KARRAS et al., 2019a). Os autores deixam claro que não modificaram o discriminador ou as funções de perda tradicionais.

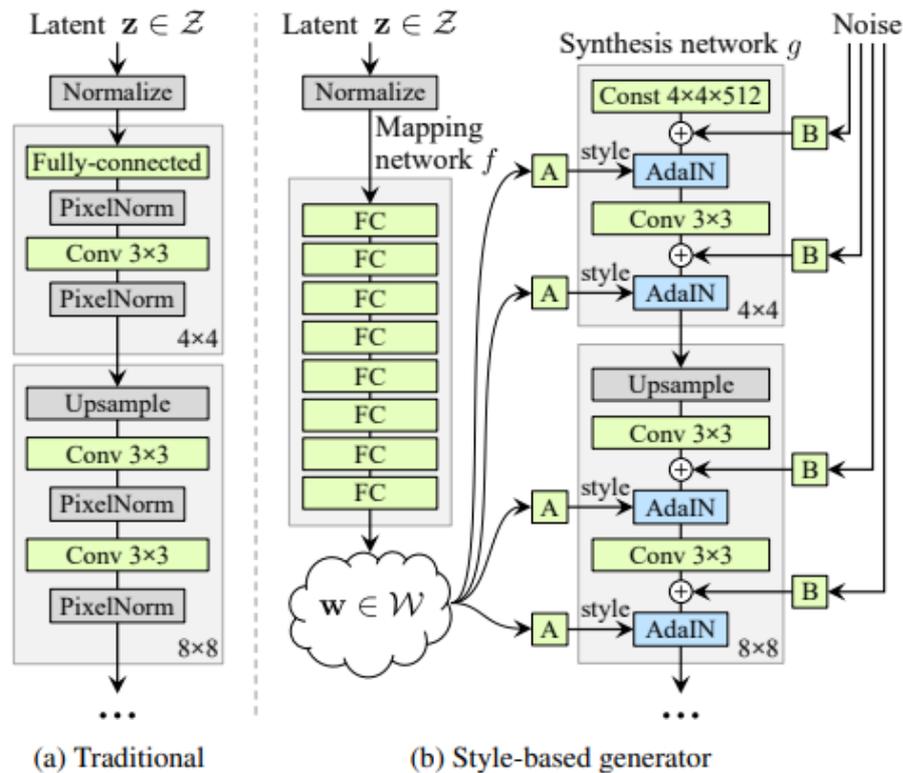
Com relação às adaptações feitas no gerador, foi torná-lo capaz de incorporar o código latente de entrada em um espaço latente intermediário, que tem um efeito significativo em como os vetores de variação são representados na rede. Segundo Karras et al. (2019a), o espaço latente de entrada deve seguir

a densidade de probabilidade dos dados de treinamento e isso leva algum grau de emaranhamento inevitável. Diferente disto, o espaço latente intermediário de Karras et al. (2019a) está livre dessa restrição, podendo ser desemaranhado.

O dataset utilizado por Karras et al. (2019a) foram dois, sendo: CelebA-HQ (KARRAS et al., 2018) e Flickr Faces HQ (FFHQ). O CelebA-HQ é um dataset público, uma versão do CelebA criada e apresentada por Karras et al. (2018) com 30 mil imagens faciais de alta qualidade com resolução de 1024 x 1024. Já o FFHQ, foi um novo dataset apresentado por Karras et al. (2019a), consistindo o tamanho de 70.000 imagens de alta qualidade com resolução de 1024 x 1024. Segundo os autores, o FFHQ inclui uma variação muito maior do que o CelebA-HQ em termos de idade, etnia, ambiente de fundo da imagem e acessórios como óculos, óculos de sol, chapéus, dentre outros. Karras et al. (2019a) disponibilizaram o dataset FFHQ publicamente no repositório Github⁵ oficial da pesquisa.

A respeito da arquitetura do gerador proposta, diferente do gerador tradicional (Figura 18-a) em que alimenta o código latente através da camada de entrada apenas, no gerador proposto Style-based antes de tudo é mapeado a entrada Z para um espaço latente intermediário W , que controla o gerador em cada camada de convolução por meio de um processo chamado *adaptive instance normalization (AdaIN)* ou normalização adaptativa de instância (KARRAS et al., 2019a).

Figura 18 – Arquitetura de um gerador tradicional (a), e arquitetura do gerador baseado em estilo proposto pelos autores (b).



Fonte: (KARRAS et al., 2019a)

Com relação ao mapeamento, o processo pode se dar da seguinte forma: dado um código latente z no espaço latente de entrada Z , é mapeado por uma rede de mapeamento não linear, podendo ser representado como $f : Z \rightarrow W$ primeiro irá produzir $w \in W$. Em seguida para fins de simplificação, os autores definiram a dimensionalidade de ambos os espaços latentes em 512, e o mapeamento f é implementado usando 8 camadas MLP totalmente conectadas (KARRAS et al., 2019a).

Em seguida, após mapear a entrada para um espaço latente intermediário, o ruído gaussiano é adicionado após cada convolução, antes de avaliar a não linearidade. Na Figura 18-b, **A** representa uma transformação afim aprendida e em **B** aplica fatores de escala por canal aprendida à entrada de ruído.

A respeito das redes que compõem a arquitetura, como mencionado

acima, a rede de mapeamento consiste em 8 camadas totalmente conectadas e a dimensionalidade de todas as ativações de entrada e saída, incluindo z e w é 512. Já a rede de síntese g , consiste em 18 camadas, duas para cada resolução (4×4 - 1024×1024). A saída da última camada é convertida em RGB usando uma convolução 1×1 separada (KARRAS et al., 2019a).

Por fim, a função de ativação usada por Karras et al. (2019a) foi a leaky ReLU com $a = 0,2$ e taxa de aprendizagem equalizada para todas as camadas. Os pesquisadores também afirmam que o gerador proposto tem um total de 26,2 milhões de parâmetros treináveis, isso em comparação com 23,1 milhões do gerador tradicional (KARRAS et al., 2019a).

Para avaliação do modelo, Karras et al. (2019a) propõe duas novas métricas automatizadas, sendo: comprimento do caminho perceptivo e separabilidade linear. Isso, para quantificar os aspectos do gerador, como o emaranhamento. Usando estas métricas em comparação com uma arquitetura de gerador tradicional, os autores puderam notar que o gerador desenvolvido por eles admite uma representação mais linear e menos emaranhada de diferentes fatores de variação.

Os resultados obtidos pela métrica do comprimento do caminho perceptivo, estão representados pela Tabela 5, que mostra que o comprimento de caminho completo é substancialmente mais curto para o gerador baseado em estilo com entradas de ruído desenvolvido pelos pesquisadores. Essa informação relatada pode ser observada em negrito pela coluna **Comprimento do caminho**, onde tanto o completo quanto o final são mais curtos, indicando que W é perceptualmente mais linear do que Z (KARRAS et al., 2019a).

Tabela 5 – Comprimentos de caminho de percepção e pontuações de separabilidade para várias arquiteturas de gerador em FFHQ (quanto menor, melhor).

Método	Comprimento do caminho		Separabilidade
	Completo	Final	
B Gerador tradicional Z	412.0	415.3	10.78
D Gerador Style-based W	446.2	376.6	3.61
E + Adição de entrada de ruídos W	200.5	160.6	3.54
+ Mistura 50% W	231.5	182.1	3.51
F + Mistura 90% W	234.0	195.9	3.79

Fonte: (KARRAS et al., 2019a)

A Tabela 6, mostra os resultados de como os comprimentos de caminho são afetados pela rede de mapeamento. Nos resultados apresentados na tabela, é perceptível que em ambos os geradores, tanto o tradicional quanto o baseado em estilo, se beneficiam de uma rede de mapeamento, e a profundidade adicional geralmente melhora o comprimento do caminho perceptivo, bem como os *Fréchet inception distances (FIDs)* que quanto menor, melhor. De forma geral, com base nos dados apresentados, uma rede de mapeamento mais profunda geralmente tem um desempenho melhor do que uma superficial (KARRAS et al., 2019a).

Tabela 6 – O efeito de uma rede de mapeamento em FFHQ. O número no nome do método indica a profundidade da rede de mapeamento.

Método	FID	Comprimento do caminho		Separabilidade
		Completo	Final	
B Tradicional 0 Z	5.25	412.0	415.3	10.78
Tradicional 8 Z	4.87	896.2	902.0	170.29
Tradicional 8 W	4.87	324.5	212.2	6.52
Style-based 0 Z	5.06	283.5	285.5	9.88
Style-based 0 Z	4.60	219.9	209.4	6.81
Style-based 0 Z	4.43	217.8	199.9	6.25
F Style-based 8 W	4.40	234.0	195.9	3.79

Fonte: (KARRAS et al., 2019a)

Já os resultados obtidos através da separabilidade linear, também podem ser observados, pelas Tabelas 5 e 6 que mostram na coluna **Separabilidade** que *W* é melhor separável do que *Z*, desta forma sugerindo uma representação menos emaranhada. Foi observado que ao aumentar a profundidade da rede

de mapeamento melhora a qualidade da imagem e a separabilidade linear em W . Já ao adicionar uma rede de mapeamento na frente de um gerador tradicional resulta em uma severa perda de separabilidade em Z , como pode ser observado pela coluna **Separabilidade** na segunda linha da Tabela 6, mas em contrapartida melhora o FID e a separabilidade no espaço latente intermediário W (KARRAS et al., 2019a).

3.5 Correlações entre os trabalhos e a pesquisa

Analisando os trabalhos apresentados neste capítulo, pontua-se que nos três primeiros (ANTIPOV et al., 2017; LIU et al., 2019; SHEN et al., 2020), em que o objetivo se assemelha bastante ao que se pretende realizar aqui. O quarto trabalho, de Karras et al. (2019a), ajudou bastante para a compreensão da arquitetura, uma vez que, usaremos esta arquitetura como base para a solução proposta. Além de tudo, com relação aos três primeiros, foram de grande proveito para auxiliar na escolha do método que será empregado, bem como o dataset e os tipos de problemas ou desafios que poderão ser encontrados e que devemos nos atentar.

Na Tabela 7 realizamos uma comparação dos dados coletados dos trabalhos correlatos. As informações presentes nesta tabela, foram importantes para o planejamento e execução deste TCC. A coluna T.GAN da Tabela 7 correspondente aos tipos de GAN, nela são apresentadas as estruturas utilizadas nos trabalhos. Baseado nesta primeira coluna da tabela e nos estudos dos trabalhos correlatos, inicialmente a estrutura de GAN que se adotou neste trabalho são de GANs baseadas na literatura de transferência de estilo, às StyleGANs, visto que estas apresentaram resultados significativos.

Na Tabela 7, as colunas Dataset, Disp. Dataset e Tam.Dataset correspondem respectivamente aos bancos de dados, disponibilidade do banco de dados, e ao tamanho do banco de dados. Na Tabela 7, a coluna Dim. Imagens, refere-se a dimensão das imagens contidas nos datasets. As informações obti-

das a respeito das dimensões das imagens, auxiliam bastante a compreender a necessidade de cada arquitetura específica. Como pôde ser observado que os datasets preparados para GANs tradicionais geralmente possuem dimensões de imagens inferiores a 256 x 256, já para as GANs de alta qualidade normalmente as dimensões são de 1024 x 1024. Essas são informações que nos levam a uma atenção maior para caso surja a necessidade de um pré-processamento das imagens, seja, redimensionando, recortando ou aumentando o nível de nitidez.

Por último a coluna Result. de acurácia, na Tabela 7, corresponde aos resultados que os trabalhos correlatos obtiveram na classificação. Os resultados obtidos pelos trabalhos correlatos, são resultados consideráveis, o que mostra a estabilidade do modelo proposto pelos autores de todos estes trabalhos. Fazendo com que se tenha uma segurança maior ao se basear nestes trabalhos.

Tabela 7 – Classificação dos artigos por técnicas.

Artigos	Dados dos Trabalhos Correlatos					
	T.GAN	Dataset	Disp. Dataset	Tam. Dataset	Dim. Imgens	Result. de acurácia
Trabalho 1	Age-cGAN	IMDB-Wiki_cleaned	Público	120K	Não consta	82.9%
Trabalho 2	Wavelet-based GAN	MORPH e CACD	Privado	55K e 163K446	Não consta	98.26% e 98.44%
Trabalho 3	StyleGAN e PGGAN	CelebA e CelebA-HQ	Público	200k+ e 30K	178 x 218 e 1024 x 1024	97.9% e 75.3%
Trabalho 4	StyleGAN	CelebA-HQ e FFHQ	Público	30K e 70k	1024 x 1024	–

Fonte: Própria do autor.

4 Método da Solução Proposta

Este capítulo apresenta a solução proposta, definindo sua arquitetura e fluxo de execução. Adicionalmente, é apresentada a implementação da solução proposta em software, identificando as ferramentas e frameworks utilizados.

4.1 Arquitetura do Método Proposto

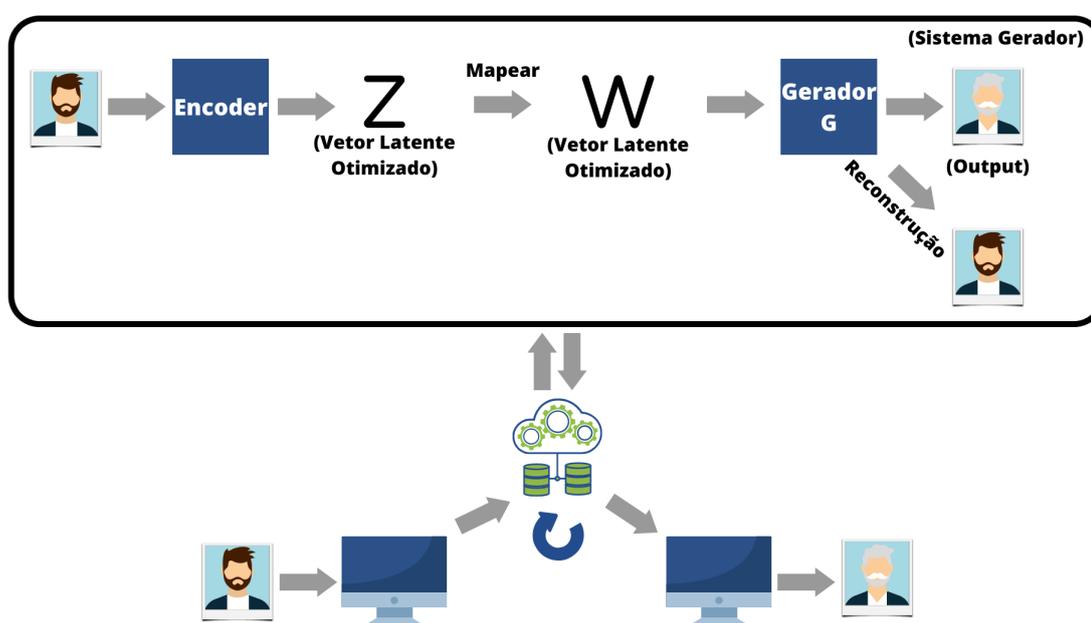
A Figura 19 apresenta uma visão geral da solução proposta que visa o desenvolvimento de um sistema de StyleGAN para geração de imagens baseado em previsão temporal para pessoas desaparecidas. Inicialmente realiza-se a entrada da foto de uma pessoa, a qual se deseja prever seu envelhecimento. Em seguida, de acordo com a Figura 19, a foto é passada para o sistema de processamento (StyleGAN) que está representado como uma nuvem. Após o processamento da imagem, será obtido como resultado, várias imagens já envelhecidas dentro de uma determinada faixa de idade, o fator determinante para a faixa etária são os parâmetros fornecidos no software `--steps` juntamente com `--start_distance` e `--end_distance` que são explicados durante a Seção 4.3.

Para a execução do sistema, foi utilizado modelos pré-treinados, que foram disponibilizados pelos autores oficiais Karras et al. (2019a) e Shen et al. (2020) em seus repositórios Github^{5,6}. Os modelos pré-treinados, foram treinados usando as bases de dados públicas CelebA-HQ (KARRAS et al., 2018) e FFHQ (KARRAS et al., 2019a), uma vez que segundo o autor Karras et al. (2019a) o dataset FFHQ em relação ao CelebA-HQ possui uma grande variabilidade em termos de idade, etnia, acessórios faciais. Tendo este diferencial entre os datasets, a avaliação do sistema gerador (representado pelo retângulo preto na Figura 19) de uma maneira mais dinâmica para coletar resultados.

A próxima etapa da solução proposta que é indicada pelo símbolo da

nuvem na Figura 19, a imagem dada como entrada é passada para o encoder que fica encarregado por gerar um vetor de ruído Z otimizado (também podendo ser chamado de vetor latente) no espaço latente Z , em seguida, este vetor é mapeado para um espaço latente intermediário W e em seguida passado para o gerador G , que por sua vez gera a uma reconstrução inicial da representação do vetor latente e resulta nas imagens envelhecidas dentro de uma faixa de etária.

Figura 19 – Funcionamento do sistema proposto.

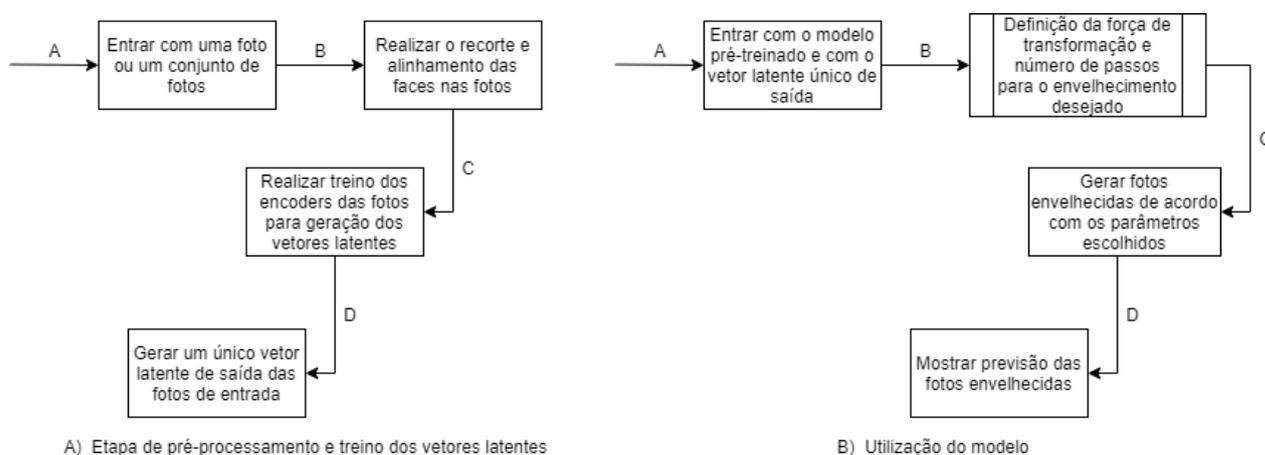


Fonte: Própria do autor.

A Figura 20 apresenta dois diagramas de fluxo, onde o primeiro (Figura 20-A) apresenta o fluxo de execução da etapa de pré-processamento e treino dos vetores latentes para uso do modelo da solução proposta e o segundo (Figura 20-B) apresenta fluxo da sequência de passos da utilização do modelo. A seguir será explicado o funcionamento de cada uma das etapas dos respectivos diagramas.

Antes da utilização do modelo são necessárias algumas etapas, as quais são apresentadas na (Figura 20-A). Durante a primeira etapa (etapa A), deve-se realizar a entrada de uma foto, ou conjunto de fotos as quais se desejam

Figura 20 – Diagrama de fluxo.



Fonte: Própria do autor.

realizar a previsão temporal. em seguida, na etapa B, caso as imagens não estejam nos padrões requeridos pelo modelo, ou seja, caso elas não estejam em uma alta qualidade de 1024 x 1024 e com os rostos não centralizados, será realizado o pré-processamento (redimensionamento e/ou recorte para a centralização da face) destas imagens. Já na etapa C, a partir das imagens obtidas pelo pré-processamento, será realizado o treino dos encoders com as fotos já pré-processadas para a obtenção dos vetores latentes de cada uma. Por fim, na etapa D, deve-se gerar um único vetor latente com os vetores de saída obtidos no passo anterior para cada foto.

Já as etapas para a utilização do modelo, são apresentadas pela (Figura 20-B). Na primeira etapa (etapa A), deve-se realizar a entrada de um modelo pré-treinado e do vetor latente de saída único obtido pela fase anterior. Em seguida, na etapa B, deverá ser definido os parâmetros para o envelhecimento desejado, como a força de transformação (`--start_distance` e `--end_distance`) e o número de passos para que as transformações sejam aplicadas. Feito isso, na etapa C, é o momento em que o modelo gerador fica responsável por gerar as previsões de acordo com os parâmetros escolhidos. Por fim, na etapa D, é o momento em que, respectivamente, serão mostrados os resultados das reconstruções iniciais dos vetores latentes e os resultados

das previsões de envelhecimento.

4.2 Ferramentas e Implementações

A solução proposta deste trabalho foi implementada em código escrito na linguagem Python ¹(v3.7.12) usando principalmente os *frameworks* Tensorflow ²(v2.4.1), que foi necessário utilizar o modo de compatibilidade com o tensorflow v1.x, e Keras ³(v2.3.1) que é uma API que rodará em cima do Tensorflow. Estes *frameworks* serão utilizados em conjunto em grande parte da construção do modelo, principalmente serão para a implementação do modelo perceptual do *Encoder*, do Gerador e do Discriminador. Desta forma, possibilitando a geração dos vetores para uso da aplicação de previsão. A função principal do Keras API no modelo é prover suporte a implementação do modelo de Karras et al. (2019a), e auxiliar nas etapas de construção de todas as redes que irão compor o modelo como a rede perceptual e geradora do *encoder*, que foi uma contribuição da comunidade ⁴ ao trabalho de Karras et al. (2019a).

4.3 Estrutura do GAN-AGE e o Fluxo de Execução

A solução proposta é dividida em duas partes, sendo a primeira para o treino dos encoders, onde foi adotado a solução que utiliza styleGAN, encontrada no repositório do GitHub mantido pela comunidade ⁴, esta solução foi inicialmente proposta como contribuição ao repositório oficial dos pesquisadores Karras et al. (2019a) ⁵. Esta parte é responsável por transformar as imagens em uma representação de código latente, ou seja, possui a finalidade de obter os vetores latentes. A segunda parte da solução, que consiste na predição, foi adotado a ferramenta InterFaceGAN disponível no repositório oficial de Shen et

¹ <https://www.python.org>

² <https://www.tensorflow.org>

³ <https://github.com/keras-team/keras>

⁴ <https://github.com/pbaylies/stylegan-encoder>

⁵ <https://github.com/NVlabs/stylegan>

al. (2020) ⁶, sendo esta parte responsável por interpretar os vetores latentes das GANs e gerar as imagens previstas.

A aplicação da styleGAN proposta por Karras et al. (2019a), tem sido utilizada como base e inspiração para muitas pessoas da comunidade, já contando com mais de 68 implementações diferentes, incluindo as implementações oficiais. Já a aplicação InterfaceGAN proposta por Shen et al. (2020), apesar de ser relativamente recente, já conta com diferentes contribuições efetuadas pela comunidade no repositório do GitHub.

Na aplicação do InterfaceGAN foi adotado alguns modelos pré-treinados para serem usados em conjunto com algum dos limites pré-definidos, que nada mais são do que o limiar/fronteira de mudança das características faciais. Alguns exemplos de modelos pré-definidos são: os limites de idade para o espaço latente Z e W , de idade condicionado ao gênero ou a um acessório facial (exemplo, óculos). Ao todo são 39 limites pré-treinados, disponíveis para cada caso de uso. Já os modelos pré-treinados foram disponibilizados pelos autores oficiais da InterfaceGAN, esses modelos são então treinados usando uma StyleGAN em cima dos datasets CelebA-HQ (KARRAS et al., 2018) e FFHQ (KARRAS et al., 2019a). Outro possível modelo pré-treinado é o de uma PAGGAN em cima do dataset CelebA-HQ. O modelo treinado adotado para este trabalho foi o modelo da StyleGAN treinado no dataset FFHQ.

Nos modelos pré-treinados do InterfaceGAN, em cada dataset, são dispostos alguns tipos de limites também pré-treinados no formato de vetores latentes, no espaço latente Z ou W . Os limites disponíveis para uso, se referem ao que se deseja trabalhar, tais como: predição de expressão facial (sorriso), idade, gênero, pose, acessório facial (óculos) e qualidade da imagem. Também estão disponíveis limites condicionados a outras informações, tais como: idade condicionada ao gênero, gênero condicionado a idade, óculos condicionado a idade, ou gênero ou ambos simultaneamente.

Vale enfatizar que para que fosse possível a execução da solução proposta, foi necessário algumas adaptações nos código das aplicações **stylegan-encoder**

⁶ <https://github.com/genforce/interfacegan>

e **InterfaceGAN**, relacionadas principalmente a migração de versão e compatibilidade do *framework* Tensorflow. Adicionalmente, foram feitas algumas adaptações em um *script* da aplicação *stylegan-encoder* (o *script* `save_latent.py`) que possui a finalidade de gerar o vetor latente único de saída ⁷, visto que a InterfaceGAN é a parte da aplicação que receberá este vetor, e que a sua representação de entrada é um array de vetores latentes, no caso o vetor único de saída obtido na *stylegan-encoder*. Essa adaptação foi feita uma vez que o *stylegan-encoder* não tinha suporte para atender ao propósito e as necessidades da aplicação InterfaceGAN. Essa alteração foi feita de forma que seguisse os requisitos arquitetados e projetados na estrutura da InterfaceGAN de Shen et al. (2020), onde admite apenas a entrada de um único vetor latente (array de vetores latentes).

Em ambas as partes da aplicação, *stylegan-encoder* e InterfaceGAN, foi utilizado um tipo de GAN chamado StyleGAN, que é uma extensão da arquitetura da GAN tradicional com modificações baseadas em transferência de estilo (*Style Transfer*). De acordo com Karras et al. (2019a), esse tipo de GAN é o método mais atual para a síntese de imagens de alta resolução, se mostrando o método mais confiável em uma variedade de conjuntos de dados.

Algumas particularidades da StyleGAN é o uso de normalização de instância adaptativa, o fato de gerar a partir de um tensor de valor fixo variáveis latentes não geradas estocasticamente como em GANs regulares. Variáveis latentes geradas estocasticamente são usadas como vetores de estilo na normalização de instância adaptativa em cada resolução após serem transformadas por uma rede *feedforward* de oito camadas. Além disso, esse estilo de GAN também emprega uma forma de regularização chamada regularização mista, que mistura duas variáveis latentes de estilo durante o treinamento (KARRAS et al., 2019a).

A primeira parte da solução, o **stylegan-encoder**, é a parte em que as imagens são mapeadas por encoders e transformadas em representações de es-

⁷ <https://neurondai.medium.com/what-does-your-face-look-like-in-the-next-few-years-gans-for-face-aging-problems-8568299adfd>

paço latente (vetores). Essa transformação se dá através do pré-processamento das fotos, onde nele, elas serão redimensionadas e as faces centralizadas, em seguida pela seleção de um modelo pré-treinado e da utilização de algumas opções de algoritmos otimizadores, como: L-BFGS Single, GGT Single, ADAM Single, ADAM Batch e ADAM Slow/Precise (AGARWAL et al., 2020). Neste caso, o algoritmo de otimização adotado neste trabalho é o ADAM Slow/Precise, tendo em vista que foi com este que foram obtidos os melhores resultados. Essas etapas descritas podem ser visualizadas na Figura 24 nos itens A, C, D.

Os vetores latentes no enconder são obtidos no espaço latente Z e em seguida são mapeados para vetores latentes intermediários do espaço latente W . Espaços latentes são alguns tipos de espaços matemáticos onde em cada um deles é possível trabalhar com vetores específicos nesses espaços. Em seguida, eles são mapeados novamente para um único vetor (*output vector*) contendo todos os blocos de representação latente de cada imagem. Esses blocos então são passados para a segunda parte da aplicação, InterFaceGAN, que a partir daí, irá interpretar os blocos de representação latente para editar um estilo de rosto da base de dados de acordo com aquela representação do bloco. A etapa descrita pode ser visualizada na Figura 24 nos itens E, F e G.

Tratando-se ainda sobre os vetores latentes, na Figura 21 podemos visualizar um exemplo da estrutura interna de um vetor latente, bem como, sua representação gráfica pela Figura 22. Como podemos perceber a sua visualização remete à imagem de um ruído, esse é um dos fatores o qual podemos chamar um vetor latente, também de vetor de ruído. Já a estrutura interna do vetor único (*output vector* gerado, pode ser visualizada pela Figura 23, e com ela podemos perceber que ele nada mais é que um array de vetores latentes/blocos de representação latente de cada imagem.

Figura 21 – Estrutura interna de um vetor latente.

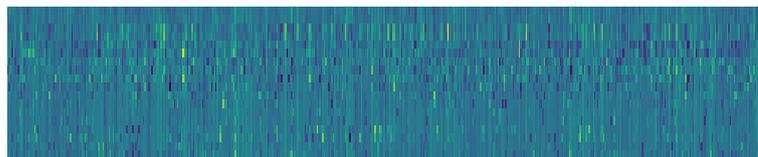
```

[[-0.11154527  0.16589175  0.05156069 ... -0.0316213  -0.00482604
 -0.1171001 ]
 [-0.06731246  0.05978122  0.00258037 ... -0.03100839  0.02400584
 -0.05867904]
 [ 0.04434378 -0.11088581 -0.11601208 ... -0.12592703  0.09350222
 -0.05128658]
 ...
 [-0.00307101  0.05989887  0.00172584 ... -0.03182965 -0.22877331
 -0.05876277]
 [ 0.04461117  0.05988777  0.00165188 ... -0.03172294  0.02381369
 -0.05854049]
 [ 0.04404766  0.1320202  -0.04795876 ... -0.03156261  0.0248781
 -0.24220091]]

```

Fonte: Própria do autor.

Figura 22 – Representação gráfica de um vetor latente.



Fonte: Própria do autor.

Figura 23 – Estrutura interna de parte de um vetor único.

```

[[[ 3.5164185e-02  8.7448388e-01  4.1190456e-03 ... -3.2328263e-02
 2.8902052e-02 -6.4246811e-02]
 [ 4.2171717e-02  6.1537929e-02  1.6383012e-03 ... -1.4362061e-01
 2.3639482e-02 -5.6101970e-02]
 [ 2.7927639e-02  5.9848741e-02 -6.9916588e-03 ... -3.6133740e-02
 2.0786921e-02 -7.0332840e-02]
 ...
 [ 8.2116053e-02  5.9428103e-02  5.1157642e-03 ... -5.8333933e-01
 -3.6734837e-01 -5.9217859e-02]
 [ 4.2782553e-02  5.9901588e-02  1.5321837e-03 ... -3.1028477e-02
 2.7191881e-02 -6.1027352e-02]
 [ 4.3684982e-02  6.0518138e-02  4.6519991e-03 ... -3.4881182e-02
 2.5748447e-02 -5.8131471e-02]]

[[ 3.8195748e-02  6.1177671e-02 -1.4239269e-03 ... -3.4417495e-02
 2.5452184e-02 -5.3888798e-02]
 [ 4.4860162e-02  1.7212820e-01 -5.4086518e-04 ... -3.8098142e-02
 4.3797788e-01 -5.9588037e-02]
 [ 2.9392673e-02  4.9178787e-02  1.3289428e-02 ... -4.0674064e-02
 2.7627770e-02 -6.1357412e-02]
 ...
 [ 3.9399948e-02  5.9380323e-02 -2.0215327e-01 ... -3.1681620e-02
 2.1798167e-02 -5.8346201e-02]
 [ 4.5449503e-02  5.9373695e-02  2.3012604e-03 ... -3.3875611e-02
 2.6413318e-02 -5.8669128e-02]
 [ 4.4052366e-02  5.7439275e-02  6.0000038e-04 ... -2.7203199e-02
 2.3466520e-02 -6.3467908e-01]]

```

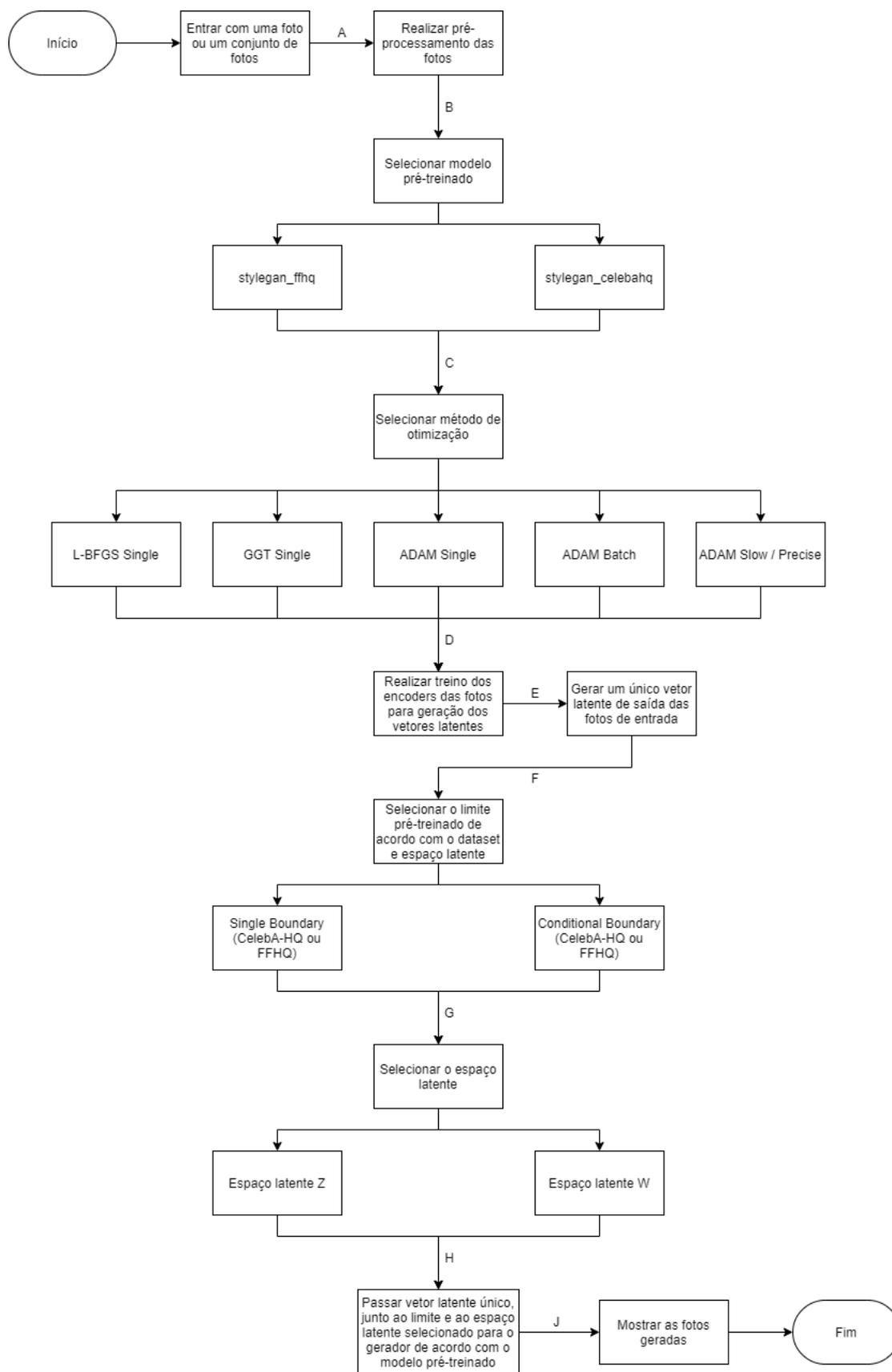
Fonte: Própria do autor.

No fluxo de execução da solução (segunda parte da solução), o **InterFaceGAN** irá realizar a predição, por meio da execução de um *script* de edição da imagem, o `edit.py`. Nele, existe alguns parâmetros necessários para execução, sendo eles:

- `-m`, para especificar o nome do modelo gerador o qual se deseja utilizar;
- `-b`, para especificar o caminho dos limites semânticos;
- `-s`, para especificar o tipo de espaço latente usado na StyleGAN;
- `-i`, para especificar o caminho dos códigos latentes da representação das imagens de entrada;
- `--start_distance` e `--end_distance`, são para especificar o intervalo de manipulação no espaço latente dos códigos latentes das imagens, ou seja, este intervalo seria a força de transformação aplicada em cada número de passos sobre a representação de código latente das imagens de entrada, ou melhor dizendo as imagens de entrada; e
- `--steps`, para especificar o número de passos para cada edição da imagem, seria em outras palavras, o número de iterações para cada mudança aplicada nas imagens.

Os parâmetros do *script* mencionado acima, são as etapas que se seguem após a etapa F da Figura 24. Tendo finalizado sua execução resta apenas a etapa J, em que nela serão mostradas as fotos geradas/preditas.

Figura 24 – Diagrama de fluxo detalhado.



Fluxo de execução da aplicação

Fonte: Própria do autor.

Na aplicação InterfaceGAN, na parte da execução do *script* `edit.py`, não fica definido qual é o melhor caso (ou seja valores) para os parâmetros que foram utilizados ou são adequados, porém é observado que alguns destes parâmetros, caso não definido um valor durante a execução, são setados valores padrões, entre o intervalo de `--start_distance` e `--end_distance`, que é definido entre -3.0 a 3.0 , e o `--steps`, que é definido como 10. Tendo executado o fluxo do método proposto, as imagens preditas estarão disponíveis para serem visualizadas e analisadas. Além do mais, neste trabalho, durante a avaliação experimental foram testados diferentes configurações, que serão discutidas no Capítulo 5, visando encontrar um conjunto que generaliza-se melhor os resultados das imagens geradas. O código da solução proposta pode ser encontrado no repositório Github⁸ deste trabalho.

Para a execução do código da solução proposta, foi adotado o ambiente Colab da Google. Assim, a execução foi definida da seguinte forma, na aplicação `stylegan-encoder`, as imagens sem processamento foram adicionadas ao diretório `raw_images` (imagens cruas). O nome dos diretórios podem ser personalizados de acordo com cada usuário. Em seguida é executado o *script* `align_images.py`, com a finalidade de pré-processar as imagens, desta forma passando como argumento o diretório das imagens sem processamento e o diretório resultante de saída para as imagens pré-processadas. Tendo obtido as imagens redimensionadas e alinhadas, o *script* `encode_images.py` é executado passando como parâmetro o diretório das imagens pré-processadas (`aligned_images`), o diretório para a reconstrução inicial das imagens (`generated_images`), o diretório para receber os vetores latentes de cada imagem (`latent_representations`) e o otimizador desejado, bem como, o número de iterações e os demais parâmetros já pré setados, um exemplo de execução dessa instrução seria o seguinte:

Listing 4.1 – Execução da solução proposta

```
1 python encode_images.py --optimizer=adam --lr=0.002 --decay_rate=0.95 \<\  
2   --decay_steps=6 --use_l1_penalty=0.3 --face_mask=True \<\  
   
```

⁸ <https://github.com/jpparreirap/TCC2>

```
3 --iterations=500 --early_stopping=False \\
4 --early_stopping_threshold=0.05 --average_best_loss=0.5 \\
5 --use_lpips_loss=0 --use_discriminator_loss=0 --output_video=True \\
6 aligned_images/ generated_images/ latent_representations/
```

Em seguida, deverá ser executado o *script* `save_latent.py` para salvar as representações latentes obtidas de cada imagem no passo anterior em um único vetor latente. Tendo feito esses passos, a tarefa na aplicação `stylegan-encoder` terá sido finalizada. Logo, adentrando o diretório da aplicação `InterfaceGAN`, deve-se executar o *script* `edit.py` passando como argumento o nome do modelo gerador, o caminho dos limites semânticos, o tipo de espaço latente, o caminho do vetor latente único, o diretório de saída para as imagens preditas, o intervalo para a força de transformação das imagens e o número de passos que devem ser aplicadas as transformações. Assim, obteremos como resultado as imagens preditas em intervalos, de acordo com o número de passos.

5 Avaliação Experimental

Este capítulo tem como objetivo descrever a execução da avaliação experimental da solução proposta neste trabalho. Durante o capítulo será apresentado o projeto e a avaliação experimental.

5.1 Planejamento e Projeto da Avaliação Experimental

O estudo foi conduzido aplicando o método proposto sobre *benchmarks* públicos pré-treinados como o FFHQ, CelebA-HQ durante a fase experimental. Os experimentos foram conduzidos a partir da plataforma de desenvolvimento Google Colab ¹ com a virtualização de uma máquina com uma GPU disponível escolhida de acordo com disponibilidade do momento, 12.72GB RAM e 68.40GB de Disco temporário (quando usado uma GPU). Com relação a disponibilidade das GPUs no Colab, incluem a Nvidia K80s, T4s, P4s e P100s. Somente é permitido escolher a GPU a qual se deseja trabalhar na versão Colab Pro (GOOGLE-COLAB, 2021).

Esta avaliação empírica tem como objetivo analisar a capacidade do método proposto em contribuir com a solução proposta neste trabalho. Desta forma, nesta avaliação, investiga-se as seguintes Questões de Pesquisa (QP):

QP1 : O sistema proposto é capaz de gerar uma imagem de um rosto humano envelhecido?

QP2 : As métricas do modelo proposto são suficientes para provar a confiabilidade deste sistema?

QP3 : O sistema proposto é capaz de se aproximar do perfil previsto de uma pessoa desaparecida com precisão, contribuindo nas buscas realizadas por investigadores?

¹ <https://colab.research.google.com>

Para podermos responder a estas questões de pesquisa, foram selecionadas fotos de quatro celebridades diferentes, sendo elas: Will Smith, Eunice Baía, Daniele Suzuki e Leonardo Dicaprio. Para que desta forma avaliar a variabilidade (capacidade de generalização do modelo) e o desempenho do modelo em quatro etnias diferentes: negros, indígenas, asiáticos e brancos.

Analisou-se também os resultados de métricas disponibilizadas por Karras et al. (2019a) e Shen et al. (2020) que são autores do modelo de GAN adotado neste trabalho. Porém, também foram refeitos os testes com algumas destas métricas como a distância de início de Fréchet (do inglês *Fréchet Inception Distance - FID*), comprimento do caminho perceptivo (do inglês *Perceptual Path Length - PPL*) e separabilidade linear (do inglês *Linear Separability - LS*).

Por fim, foi aplicado um questionário de pesquisa para o público em geral, para se analisar os resultados de desempenho e qualidade das imagens geradas pelo modelo de GAN. Este mesmo questionário, foi disponibilizado nas redes sociais do autor deste trabalho (Instagram e Facebook), bem como, em grupos e contatos pessoais no WhatsApp, permitindo ao público o compartilhamento do questionário, visando desta forma atingir o maior número possível de pessoas.

5.2 Execução dos experimentos e análise dos resultados

Nesta seção são apresentados os resultados obtidos pela execução dos experimentos, procurando responder às questões de pesquisa apresentadas anteriormente para validação do sistema proposto denominado de AGE-GAN.

5.2.1 Prova de conceito

Durante a avaliação foram feitos testes alterando os parâmetros do *script encode_images.py* que é responsável por gerar os vetores latentes. Além de também testar os diferentes otimizadores disponíveis neste *script*, visando obter

melhores resultados. Porém, os testes foram sem sucesso, gerando imagens de qualidades bem inferiores, algumas até no formato de espectro. Devido a estes testes, o otimizador mantido foi o ADAM Slow/Precise e os parâmetros para o algoritmo foram mantidos os padrões.

Também foram feitos testes alterando os parâmetros do *script* `edit.py` que é responsável por gerar as modificações nas imagens, desta forma gerando a previsão. No entanto, também foi possível perceber que quanto mais distante era colocado os valores de `--start_distance` e `--end_distance`, maiores eram as alterações sofridas pelo rosto, chegando a um ponto de ser irreconhecível. Já quando os valores para estes parâmetros eram mais próximos do limite semântico, próximos de 0, os resultados tendiam a ser melhores. E isso comprovou o que foi experimentado e afirmado no trabalho de Shen et al. (2020). As alterações nos `--steps`, não causavam alterações significativas, apenas mostravam as mudanças sendo aplicadas em menos ou mais passos. Devido a estes testes, foi optado por manter os parâmetros padrões para o algoritmo.

Visando responder a primeira questão de pesquisa (**QP1**) proposta, foram realizados testes utilizando fotos pessoais do autor deste trabalho, de pessoas próximas conhecidas, e de algumas celebridades, como: Will Smith, Eunice Baía, Daniele Suzuki, e Leonardo Dicaprio. Para avaliar se o sistema é capaz de gerar uma imagem de um rosto humano envelhecido, passamos algumas amostras de fotos dessas pessoas para o sistema InterfaceGAN, a fim de obter o envelhecimento das mesmas. Um exemplo de resultado obtido pode ser melhor visualizado pela Figura 25, onde nela podemos observar a imagem de entrada para o sistema, a reconstrução inicial, o envelhecimento predito pelo sistema e a imagem original da celebridade adulta.

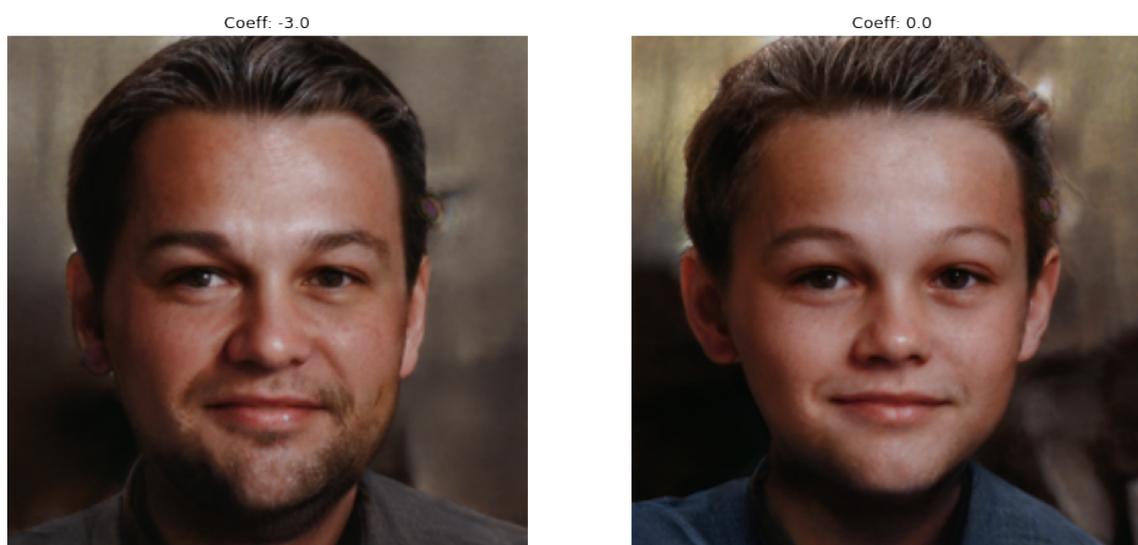
Figura 25 – Resultado de previsão da aplicação - Leonardo Dicaprio.



Fonte: Própria do autor.

Analisando os detalhes nos resultados obtidos por meio das imagens produzidas e dos coeficientes de distância percorrida, podemos perceber e comprovar, que quanto maior é a distância percorrida em relação ao limite semântico (representada pelo coef 0.0 na Figura 26), mais drásticas são as alterações sofridas pelas imagens, o que foi comentado pelo trabalho de Shen et al. (2020). Podemos visualizar essa variação da mudança nas imagens pelas Figuras 26 e 27.

Figura 26 – Forte variação dos coeficientes - Leonardo Dicaprio.



Fonte: Própria do autor.

Figura 27 – Variação detalhada dos coeficientes - Leonardo Dicaprio.



Fonte: Própria do autor.

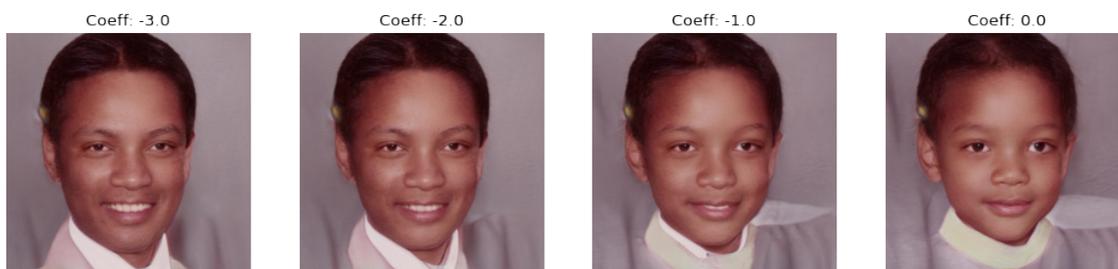
Existem casos, observados nos testes, em que a aplicação erra bastante, como foi o caso da foto da celebridade Will Smith, que pode ser vista pela Figura 28. Para este caso, o sistema não conseguiu identificar características faciais suficiente para ser caracterizado como homem durante o envelhecimento, e isso pode se dar por diversos fatores indevidos, como: a posição facial, iluminação, formato e volume da boca, pele muito jovem e lisa (não possuindo características faciais definidas), cabelo, entre outros fatores. Assim, devido a um ou mais destes fatores que podem ter influenciado nas mudanças apresentadas pela Figura 29, que mostra a variação detalhada.

Figura 28 – Resultado de previsão da aplicação - Will Smith.



Fonte: Própria do autor.

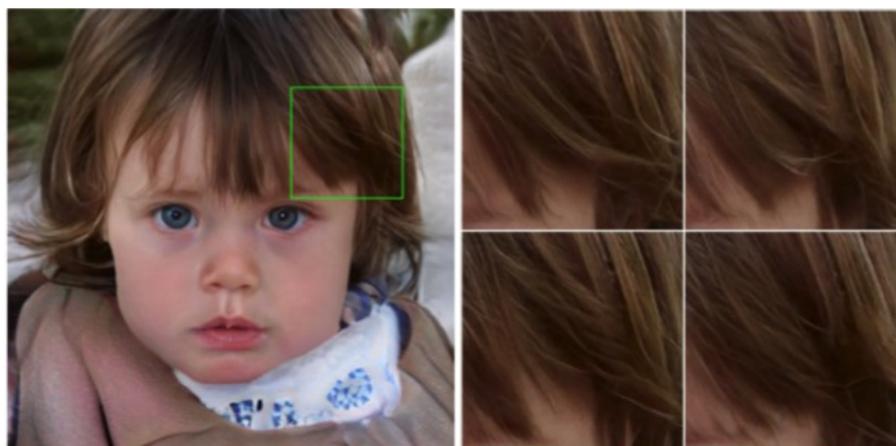
Figura 29 – Variação detalhada dos coeficientes - Will Smith.



Fonte: Própria do autor.

Ainda com relação aos erros produzidos pelo sistema, também foi possível perceber que o sistema tende a errar com a previsão de homens quando são muito jovens e não possuem características faciais definidas, como o formato do rosto um pouco mais rústico, com barba ou sinal de uma e com o cabelo mais para o lado de curto. Foi percebido que na maioria dos casos, quando o cabelo era longo o sistema tendia a transformar aquela pessoa presente na imagem em uma mulher mais velha, à medida que ia prolongando seu cabelo. Na maioria dos casos, o sistema também associava a fase idosa com a necessidade de usar óculos, e assim adicionava o acessório na face da pessoa presente na imagem. Outro ponto relevante a ser levantado é o prolongamento ou remoção dos fios de cabelo que se dão através da adição de ruído por pixel, como pode ser visto pela Figura 30.

Figura 30 – Adição de Ruído por Pixel.



Fonte: Karras et al. (2019b)

Com relação aos testes feitos com mulheres, a grande maioria deles o sistema conseguiu prever de forma eficiente, buscando seguir as características faciais detectadas pela reconstrução inicial. A seguir podemos visualizar os resultados da predição da celebridade Eunice Baía, na Figura 31, e Daniele Suzuki, na Figura 33. As Figuras 32 e 34, mostram a variação detalhada da previsão feita na Eunice Baía e Daniele Suzuki respectivamente.

Figura 31 – Resultado de previsão da aplicação - Eunice Baía.



A) Input

B) Reconstrução Inicial

C) Output

D) Original

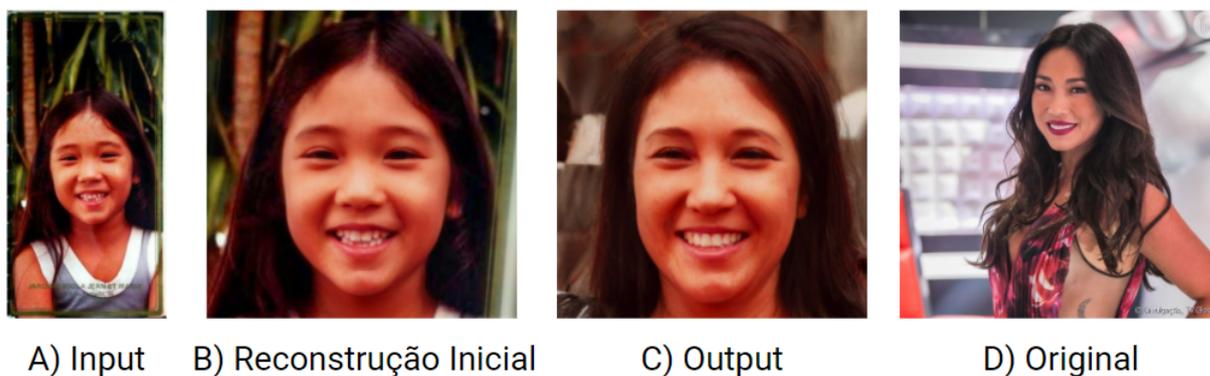
Fonte: Própria do autor.

Figura 32 – Variação detalhada dos coeficientes - Eunice Baía.



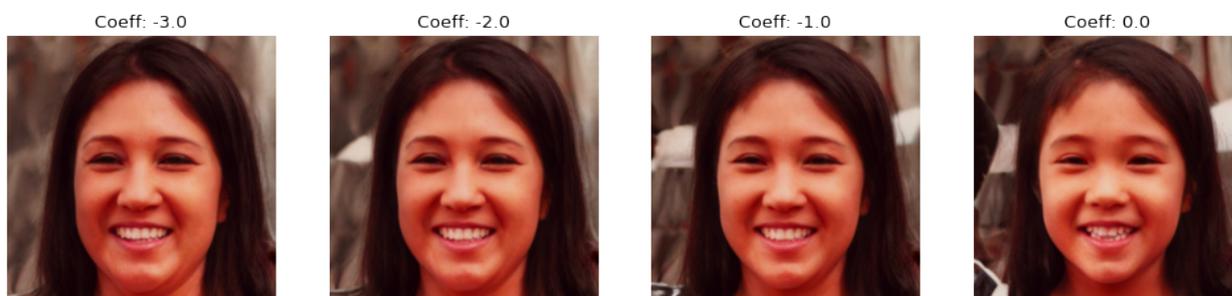
Fonte: Própria do autor.

Figura 33 – Resultado de previsão da aplicação - Daniele Suzuki.



Fonte: Própria do autor.

Figura 34 – Variação detalhada dos coeficientes - Daniele Suzuki.



Fonte: Própria do autor.

Por fim, foi feito alguns testes a fim de obter o melhor resultado de previsão, e para isso buscamos atender às condições adequadas já percebi-

das que de outra forma poderiam afetar no resultado, como: A iluminação, pose, expressão facial neutra, e características faciais definidas. Esse teste foi produzido através da foto do autor deste trabalho. Os resultados podem ser observados pela Figura 35.

Figura 35 – Resultado de previsão da aplicação - Foto Própria.



Fonte: Própria do autor.

5.2.2 Análise do modelo GAN por meio de métricas

Buscando responder a segunda questão de pesquisa (**QP2**), foram realizadas análises sob as redes do modelo por meio de métricas como distância de início de Fréchet (do inglês *Fréchet Inception Distance - FID*), comprimento do caminho perceptivo (do inglês *Perceptual Path Length - PPL*) e separabilidade linear (do inglês *Linear Separability - LS*). Além de coletar gráficos relevantes como o nível de influência da idade com relação ao gênero para a boa previsão das imagens.

Neste sentido, a métrica FID, possui a finalidade de analisar a qualidade das imagens geradas. O cálculo foi feito a partir da relação entre as imagens falsas e reais geradas. Após sua análise no modelo foi obtido o valor de 320.9395 de qualidade, avaliada em relação às imagens geradas. O resultado foi obtido a partir da avaliação de 50 mil amostras de imagens. Vale ressaltar que possivelmente os resultados seriam melhores de acordo com o número de amostras analisadas, ou seja, quanto maior o número de amostras analisadas,

melhores seriam os resultados. Essa afirmação pode ser observada através da Figura 36.

Ao aplicar a métrica PPL para o caminho completo e o final, foi possível perceber que ambos obtiveram resultados melhores no espaço latente W se comparados ao espaço latente Z. O que indica que tanto o caminho completo e final são substancialmente mais curtos em relação ao espaço Z, e além do mais, também se beneficia da rede de mapeamento, rede em que o vetor é mapeado do espaço Z para o W. O que confirma os resultados apresentados pela pesquisa de Karras et al. (2019a). Os dados podem ser observados pela Tabela 8, onde em negrito são apresentados os melhores resultados para o caminho completo e final.

Em relação à separabilidade linear, ao ser aplicada, foi possível perceber que o espaço latente W também obtém melhores resultados que o espaço Z, desta forma pode se dizer através dos resultados que o espaço W é melhor separável que o espaço Z. O que também confirma os resultados apresentados pela pesquisa Karras et al. (2019a). Os dados podem ser observados através da coluna separabilidade na Tabela 8.

Com relação a quantidade de amostras utilizadas para avaliação das métricas, foram utilizadas 50 mil amostras de imagens geradas para todas as métricas, com exceção da separabilidade linear que foram utilizadas 10 mil amostras, isso devido a um limitante de poder computacional, pois como mencionado no início deste capítulo, é uma plataforma que possui algumas limitações, o que implicou neste caso, em um longo período de tempo para a execução de 50 mil amostra e na plataforma encerrando a conexão devido ao limite máximo de tempo permitido.

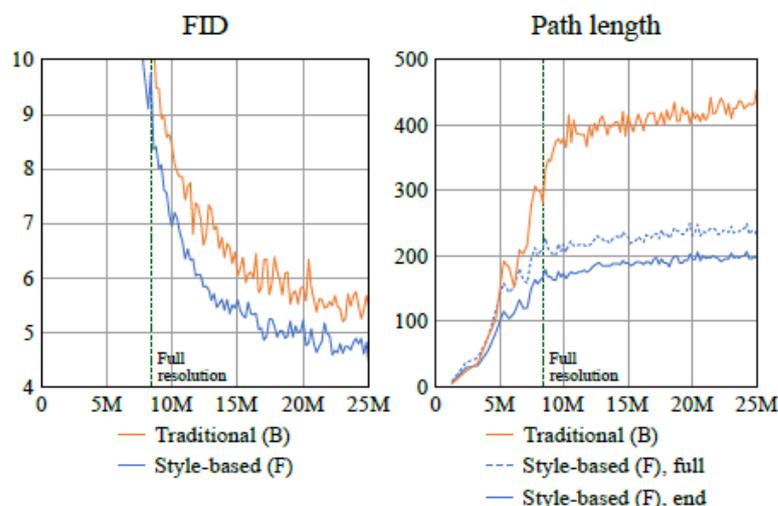
Tabela 8 – Resultados obtidos por meio das métricas PPL e separabilidade linear para cada espaço latente.

Espaço Latente	Comprimento do caminho		Separabilidade
	Completo	Final	
Z	670.7377	660.6735	44.3297
W	235.3640	197.4768	1.3831

Fonte: Própria do autor.

Outro ponto relevante para avaliação é o curso que as métricas seguem ao longo do treinamento. Na Figura 36, a linha tracejada verticalmente nos gráficos referem-se o ponto em que o treinamento progrediu as imagens para a resolução total de 1024x1024, e o eixo horizontal do gráfico refere-se a quantidade de imagens que foram vistas pelo discriminador. Como podemos perceber, o FID, tende a ser menor em GANs baseadas em estilo do que em tradicionais, como a StyleGAN, e também em relação a quantidade de imagens já vistas e aprendidas pelo discriminador. Já em relação ao PPL, GANs tradicionais tendem a aumentar muito os resultados para o caminho completo e final, obtendo assim uma piora em relação às GANs baseadas em estilo, que se mantêm mais estáveis e a um valor menor para o caminho completo e final.

Figura 36 – Curso que métricas FID e PPL seguem em relação ao treinamento usando o dataset FFHQ.

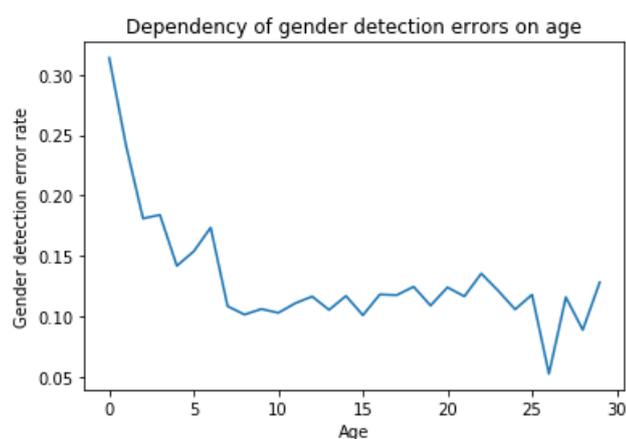


Fonte: (KARRAS et al., 2019a)

Os dados apresentados por Nikitko (2019) na Figura 37, que foram avaliados usando o mesmo modelo e o mesmo dataset (FFHQ), mostra a relação direta entre a idade de uma pessoa com a dificuldade que o modelo gerador tem de detectar o gênero do indivíduo e assim realizar a predição com precisão. Como podemos observar o modelo gerador tende a errar mais quando se tratam de crianças, e isso já era algo esperado devido aos estudos feitos. Esses erros

tendem a acontecer por muitos fatores, dentre eles o principal é o fato de que crianças normalmente não possuem características faciais bem definidas até a entrada de sua adolescência, tornando-se uma tarefa difícil para o sistema gerador distinguir entre os gêneros.

Figura 37 – Relação entre a dependência da idade da pessoa com os erros de detecção no gênero.



Fonte: (NIKITKO, 2019)

5.2.3 Análise do modelo GAN na geração de imagens do rosto humano

Durante a fase de testes, foi possível perceber que o sistema gerador não se comporta muito bem com imagens que dentre alguns fatores como qualidade da imagem, iluminação, pose, são alguns dos quais que fazem o modelo obter uma previsão indesejada. Além destes, cabelos loiros e muito claros, ou que são expostos a luz gerando um brilho, também faz com que o modelo obtenha resultados indesejados. Acessórios faciais também é outro fator determinante do sucesso na previsão, como por exemplo óculos, *piercings*, dentes de ouro, brincos, entre outros. No geral, qualquer ambiente, cabelo, ou acessório que produza brilho/reflexo tendem a atrapalhar nos resultados do sistema.

Como mencionado na seção anterior, também foi observado que crianças

normalmente não possuem uma boa previsibilidade, quanto mais nova é a criança mais difícil é para o modelo obter bons resultados. E isso acontece devido a crianças ainda não possuírem características faciais tão bem definidas para a distinção do gênero e previsão da face.

Levando estes fatores em consideração foi observado que as condições ideais para que o sistema obtivesse bons resultados, era uma foto mais ou menos tradicional a de um documento de identidade, onde a face está centralizada e olhando direto para o centro, expressão facial neutra, sem acessórios faciais e em um ambiente com iluminação suave, nem muito claro, e nem muito escuro.

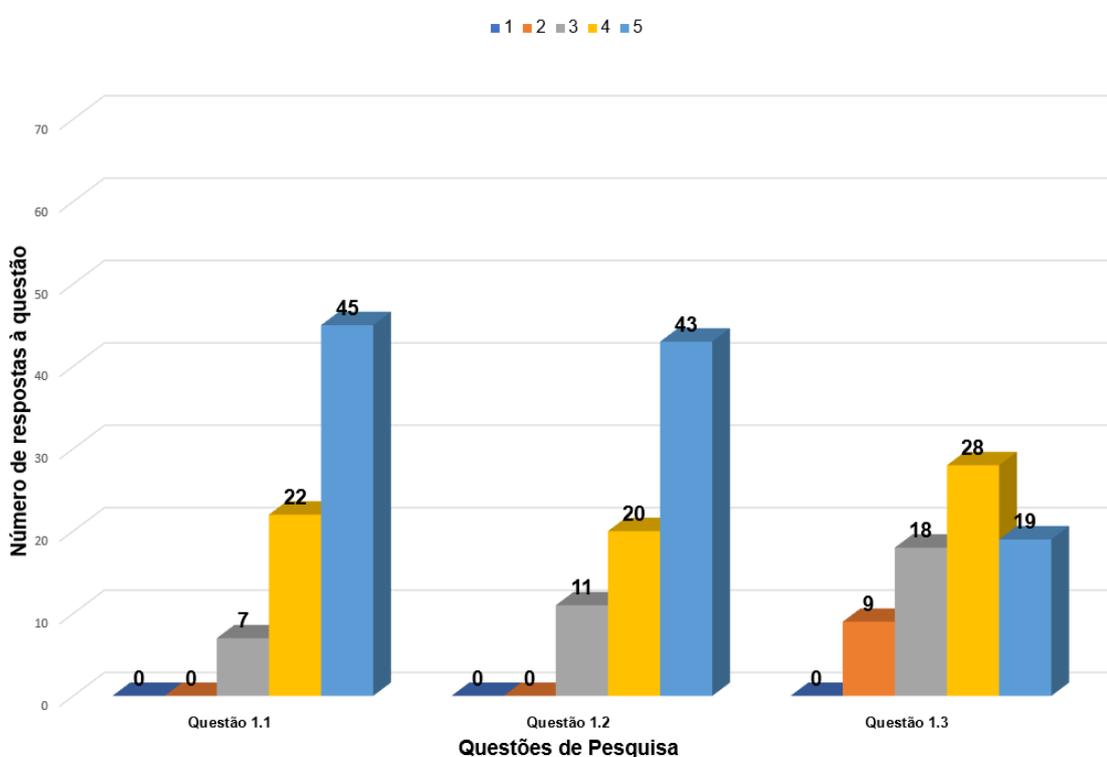
A fim de responder a terceira questão de pesquisa (**QP3**), foi aplicado um questionário para que o público pudesse avaliar os resultados da aplicação. Ao todo, foram coletadas 74 respostas que avaliaram 4 celebridades para facilitar o julgamento do público, e também para que contivesse as 4 etnias mais comuns no Brasil, isso visando avaliar também a variabilidade do modelo pré-treinado. As celebridades selecionadas foram: Will Smith, Eunice Baía, Daniele Suzuki e Leonardo Dicaprio.

Com relação a composição do questionário, ele foi dividido em 5 seções, as quais buscam avaliar: a qualidade das imagens, características faciais, qualidade do envelhecimento, grau de conhecimento/expertise de quem está respondendo ao questionário e a satisfação com resultado da solução proposta. O questionário foi projetado para atender à escala de Likert, desta forma possuindo 5 alternativas, que vão de 1 à 5, 1 para muito ruim e 5 para muito bom para a maioria das perguntas. O questionário com as questões e os resultados coletados da avaliação são apresentados no Apêndice A.

Antes de iniciar o questionário foi perguntado aos entrevistados se conheciam alguma das celebridades que seriam utilizadas como base para o questionário, e cerca de 73 pessoas responderam que sim e somente 1 pessoa respondeu que não. A maior parte do público entrevistado disse conhecer as celebridades Will Smith e Leonardo Dicaprio, assim facilitando a avaliação do resultado final da aplicação proposta.

O questionário se inicia avaliando a qualidade das imagens. Nas duas primeiras questões de pesquisa (Questões 1.1 e 1.2), buscamos avaliar respectivamente o nível de nitidez e saturação das imagens, e com base nos resultados, em que a média de respostas tende a ser acima de 40 para respostas **muito bom** (na escala sendo representado pelo valor 5), fica confirmado que o modelo gerador tende a produzir resultados muito bons para estes quesitos. Os resultados podem ser observados pela Figura 38.

Figura 38 – Respostas coletadas para a qualidade das imagens.

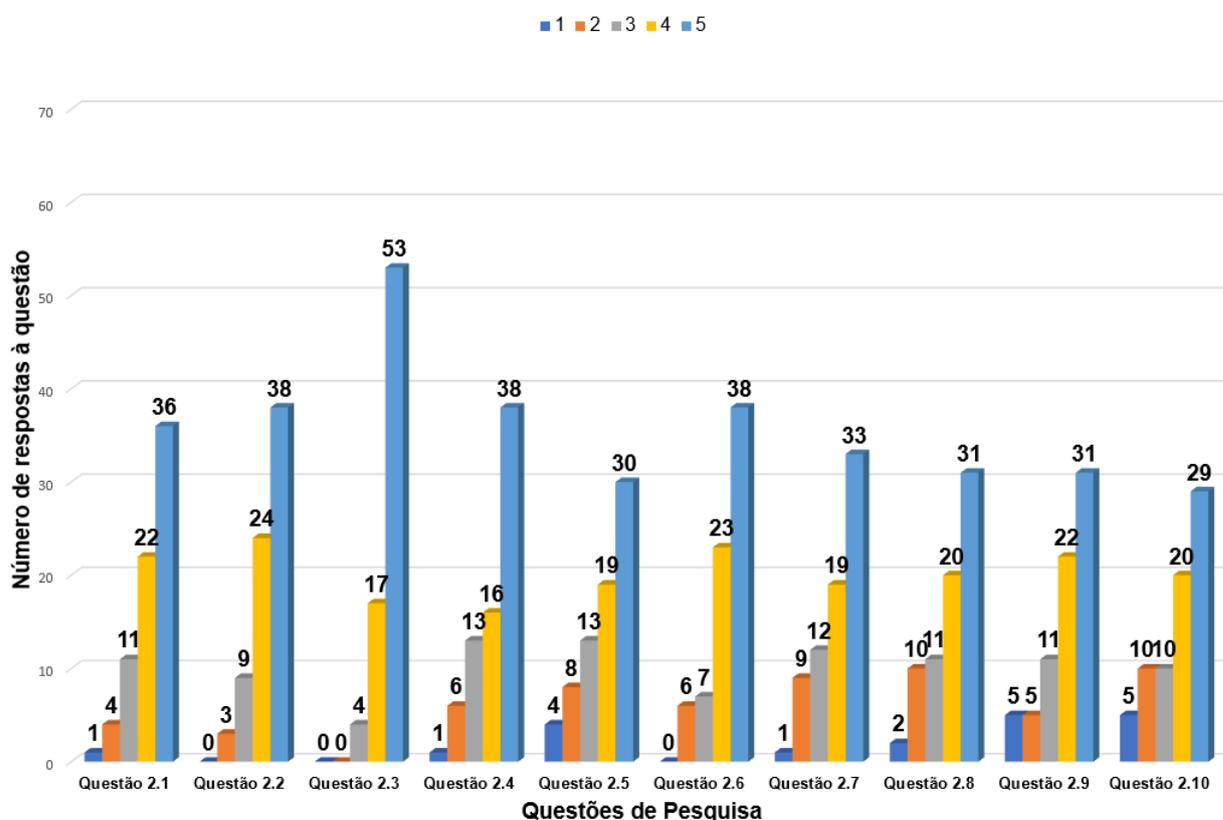


Fonte: Própria do autor.

Analisando as respostas da Questão 1.3, foi possível concluir através dos resultados que os rostos gerados possuíam a classificação de **bom** em relação aos artistas originais, além de outra grande parcela do público (cerca de 37 pessoas), se manterem divididos entre **neutro** e **muito bom**. Isso pode ser entendido devido ao fato dos resultados produzidos para a celebridade Will Smith, não serem resultados bons, precisos ou semelhantes com relação ao rosto original.

A próxima etapa do questionário é a avaliação das características faciais isoladas para cada celebridade. A primeira celebridade avaliada, foi a imagem gerada de Will Smith, os resultados podem ser observados pela Figura 39. No geral, os resultados obtidos para as características faciais se enquadraram em **muito bom**, com uma média aproximada de 36 respostas à uma classificação muito positiva e uma média aproximada de 20 à uma classificação positiva. Analisando estes resultados, observa-se uma certa estranheza, pois a previsão gerada para a celebridade em questão, visualmente foi errônea na opinião do autor deste trabalho. Porém, como o objetivo era avaliar a semelhança das características de forma isolada, o que é percebido é que algumas características obtiveram melhores resultados que outras, como por exemplo a cor dos olhos (questão 2.3) em relação aos detalhes produzidos para as orelhas (questão 2.5).

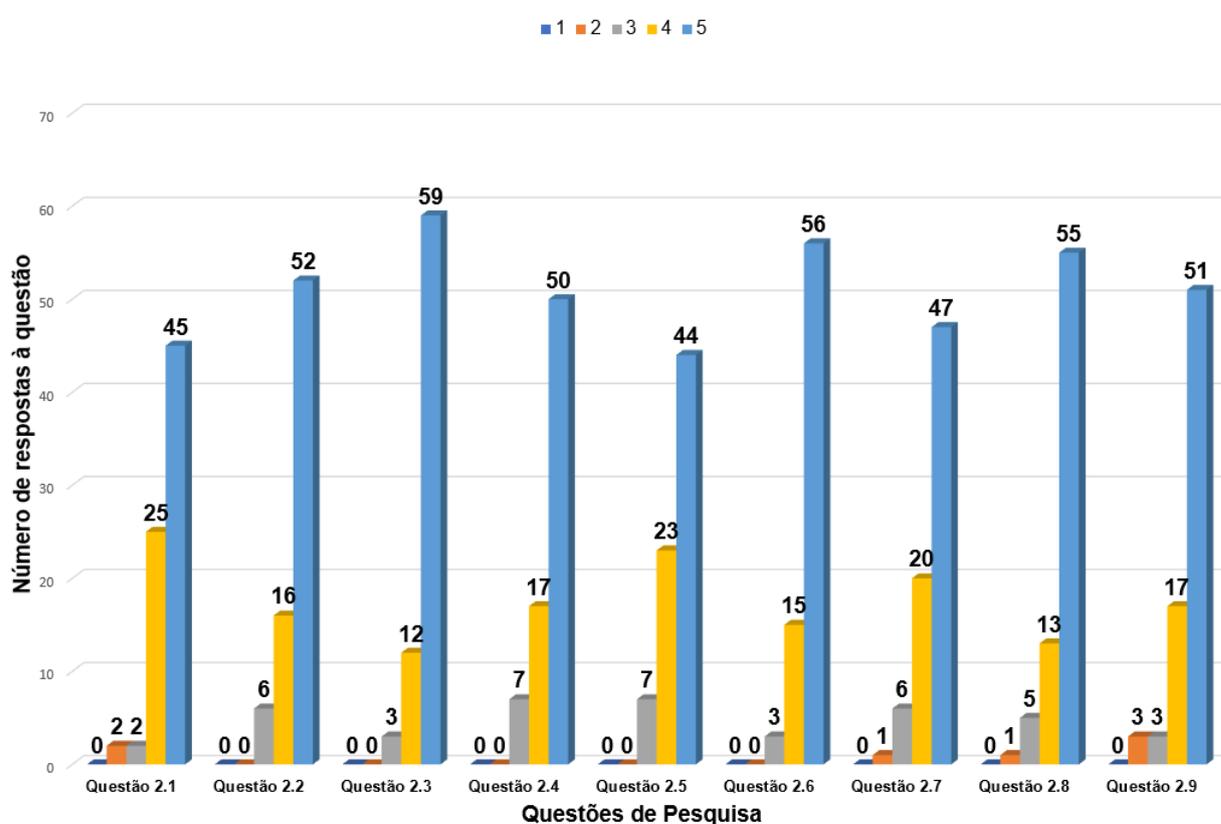
Figura 39 – Resultados para as características faciais da imagem gerada para a celebridade Will Smith.



Fonte: Própria do autor.

A segunda celebridade avaliada, foi a imagem gerada de Eunice Baía. O resultado das respostas podem ser observados na Figura 40. Os resultados obtidos na avaliação de forma geral foram extremamente positivos, ficando com uma média de 51 avaliações consideradas muito positivas. O que torna possível perceber que os resultados produzidos pelo modelo gerador para esta celebridade foram superiores e precisos para cada característica gerada.

Figura 40 – Resultados para as características faciais da imagem gerada para a celebridade Eunice Baía.

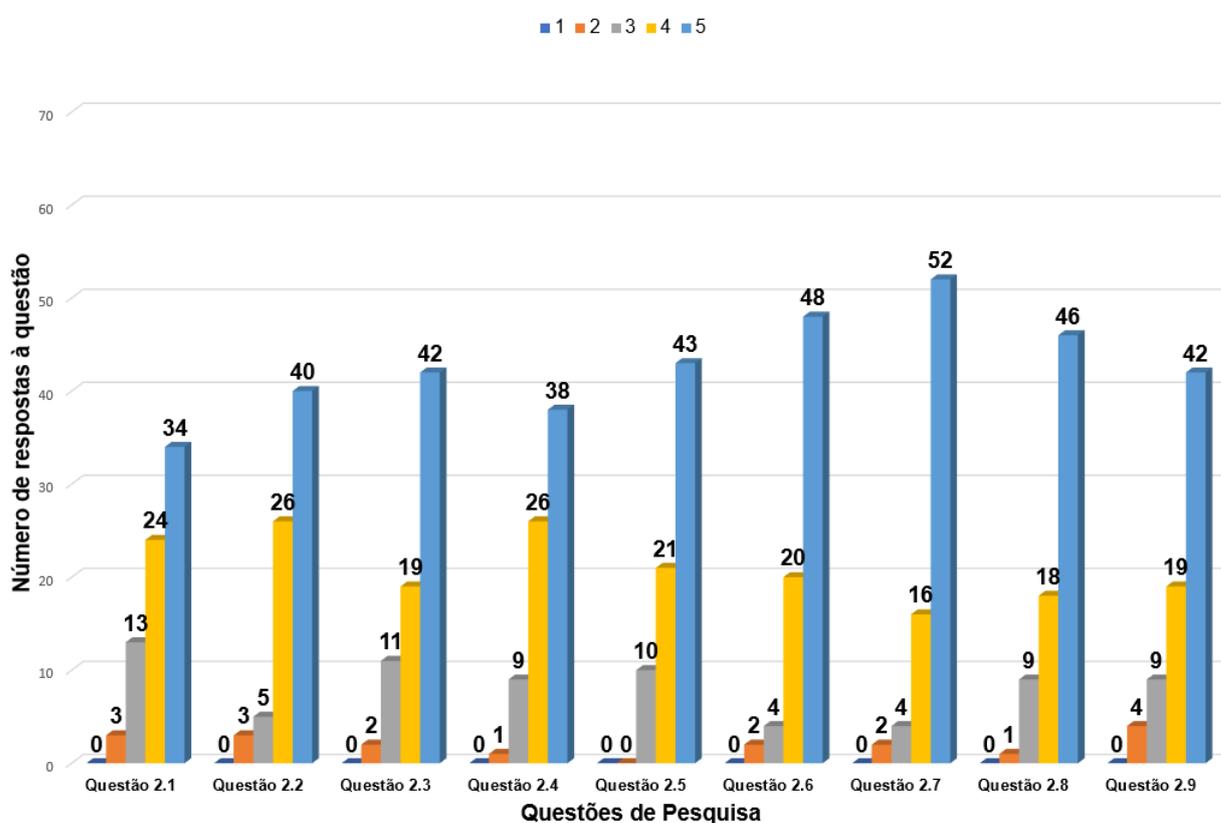


Fonte: Própria do autor.

A terceira celebridade avaliada, foi a imagem gerada de Daniele Suzuki. Os resultados obtidos na avaliação de forma geral também foram muito positivos, mas não tanto quanto para a celebridade anterior apresentada, ficando com uma média aproximada de 43 para as avaliações muito positivas. Uma possível explicação é que os resultados não tenham sido tão elevados quanto os de Eunice Baía, é o fato de a celebridade Daniele Suzuki na foto original

adulta aparentar estar com muita maquiagem, fato o qual talvez não foi levado em consideração na avaliação, uma vez que o modelo é incapaz de prever estes quesitos. Além disso, o uso de maquiagem também pode mudar drasticamente um rosto, tornando aquela pessoa diferente da aparência original, e o modelo gerador tende a seguir o envelhecimento de acordo com as características originais aprendidas na foto de entrada. Os resultados das respostas das perguntas podem ser observados na Figura 41.

Figura 41 – Resultados para as características faciais da imagem gerada para a celebridade Daniele Suzuki.

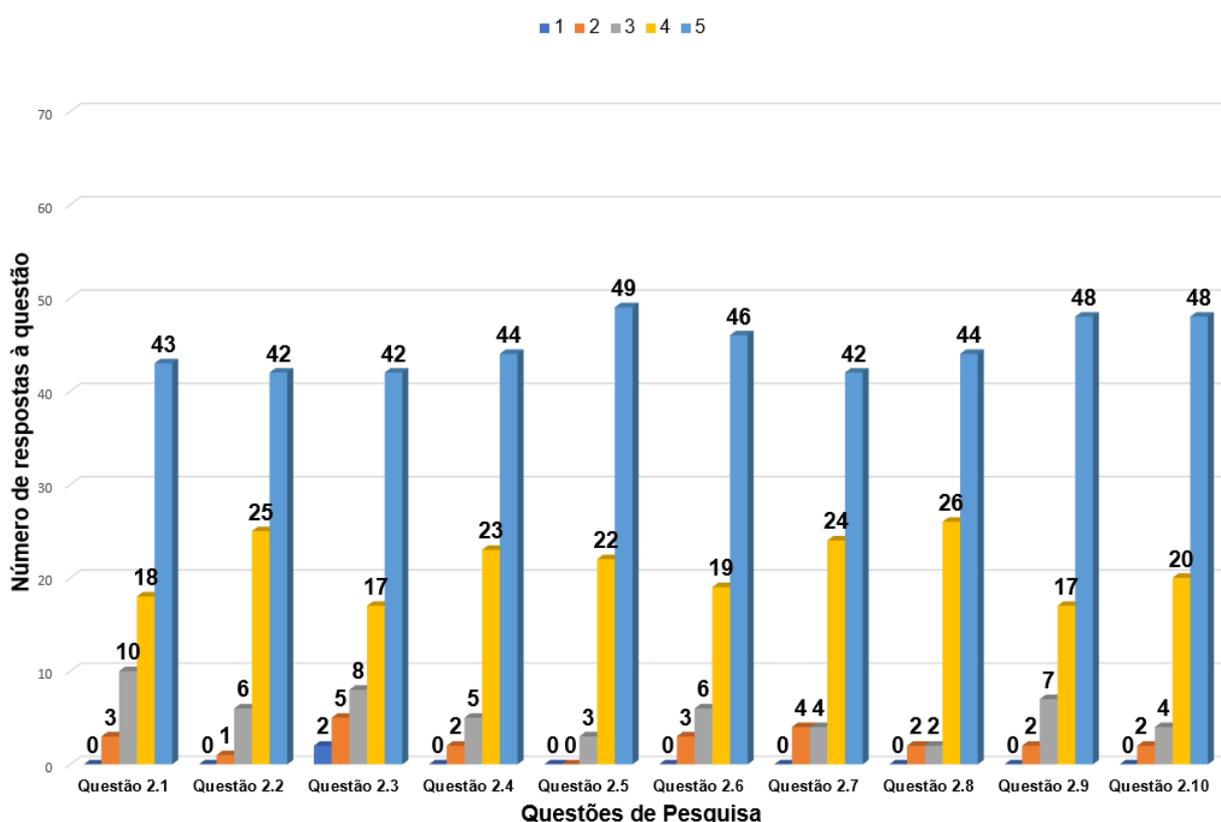


Fonte: Própria do autor.

A quarta e última celebridade avaliada, foi a imagem gerada de Leonardo Dicaprio. Os resultados obtidos na avaliação de forma geral tenderam a seguir o mesmo formato que o de Daniele Suzuki, com uma média aproximada de 45 avaliações muito positivas. Uma possível explicação para este resultado é o ambiente da foto de entrada, cujo era um ambiente escuro, o que afetou em

algumas previsões do modelo gerador, como o tom da pele e principalmente a cor dos olhos que puxaram para o lado de castanho devido a baixa iluminação que escureceu a cor dos olhos da foto de entrada. Os resultados das perguntas podem ser observados na Figura 42.

Figura 42 – Resultados para as características faciais da imagem gerada para a celebridade Leonardo Dicaprio.

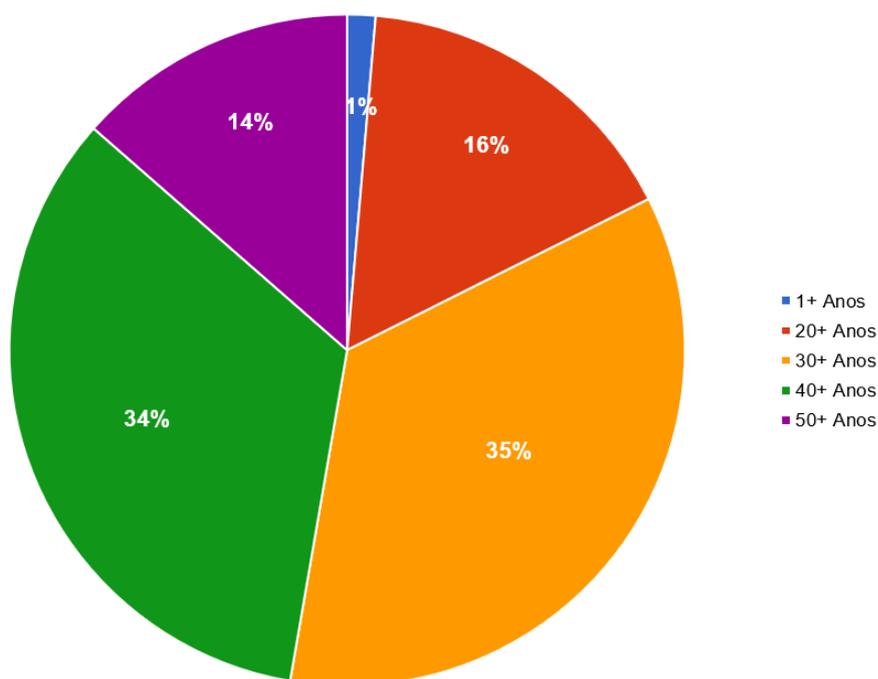


Fonte: Própria do autor.

A próxima etapa do questionário, diz respeito a qualidade do envelhecimento. Ao colocar o público para fazer o papel de um classificador (Questão 3.1), foi dada uma imagem gerada de cada celebridade e questionado a faixa etária em que se enquadravam. Considerando a Figura 43 pode-se visualizar a representação a distribuição das respostas, sendo que 69% do público ficou dividida entre mais de 30 anos e mais de 40 anos e outra parcela de 30% do público ficou dividida entre mais de 20 anos e mais de 50 anos. Foi percebido que 1 pessoa, representando 1,4%, considerou que as faces se enquadravam

em mais de 1 ano, o que isso remete a um possível dado enviesado. A classificação da maior parcela mostra que o sistema gerador consegue se aproximar da previsão da faixa etária das celebridades com precisão, uma vez que a média aproximada entre as quatro celebridades é de 44 anos.

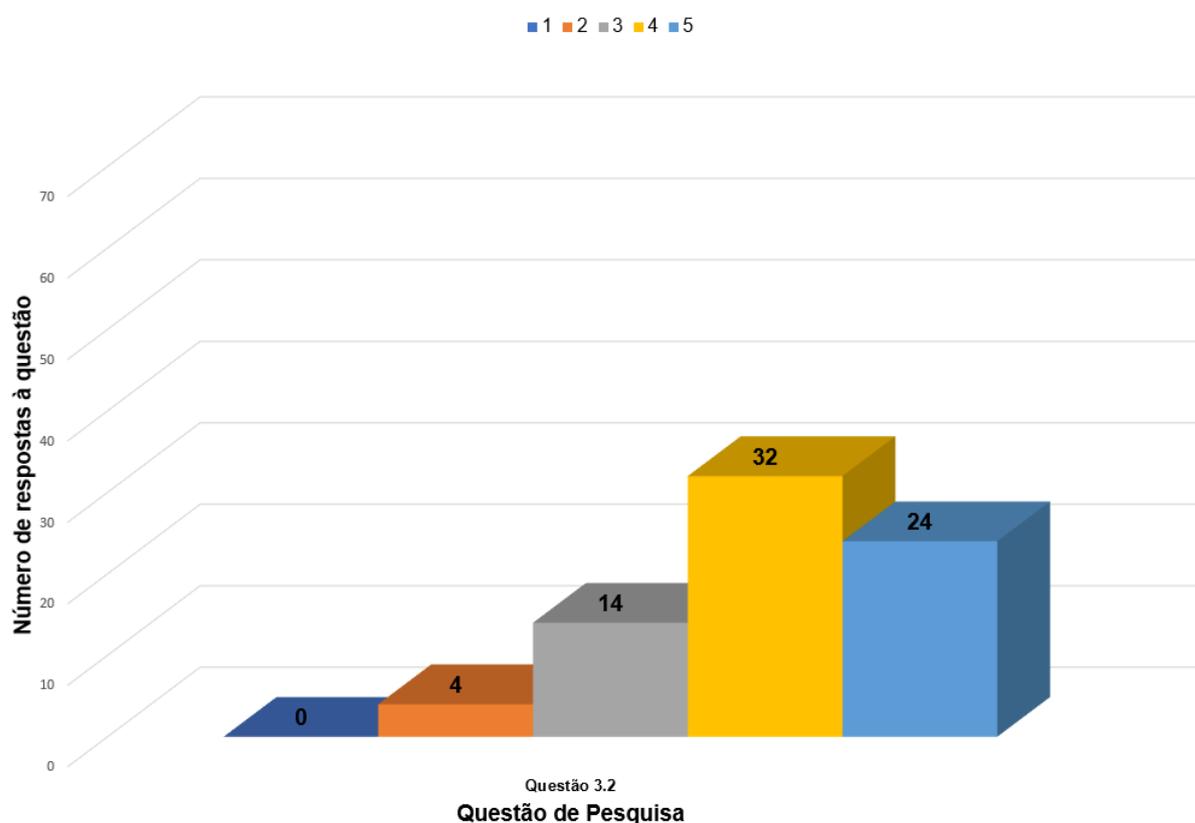
Figura 43 – Resultados para a classificação da faixa etária das celebridades geradas pelo modelo gerador.



Fonte: Própria do autor.

No questionário, quando foi perguntado ao público quanto que consideravam que as imagens obtiveram uma boa previsão de envelhecimento (questão 3.2), 32 pessoas consideraram que as imagens obtiveram uma boa previsão. Visto que as respostas ficaram divididas, isso enfatiza alguns dos fatores já mencionados, como por exemplo a previsão do Will Smith que pode ter sido afetada devido a não detecção de características definidas, também devido a imagem original da Daniele Suzuki estar com maquiagem aparente, o que tornou seu rosto diferente do rosto gerado, e a iluminação que afetou na previsão do Leonardo Dicaprio. Podemos visualizar na Figura 44 a representação gráfica dos dados.

Figura 44 – Resultados para a qualidade da previsão de envelhecimento do modelo gerador.



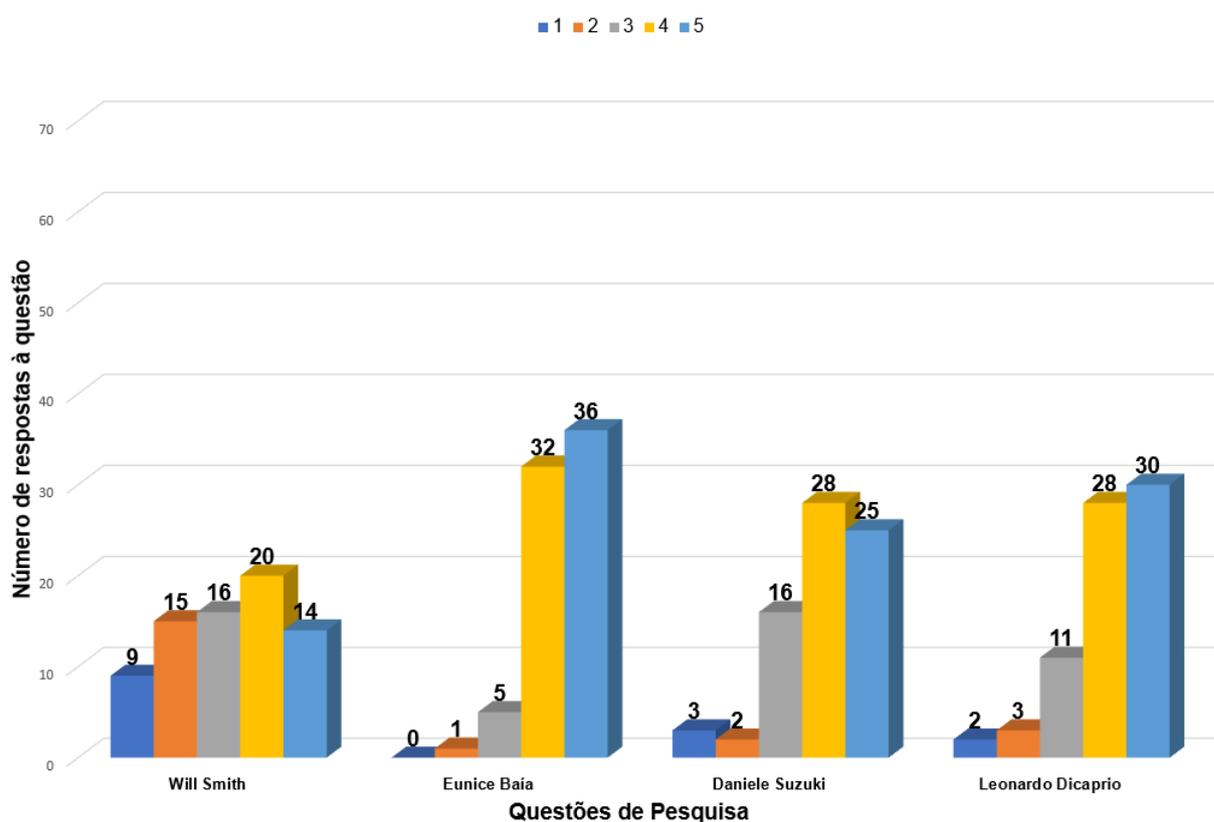
Fonte: Própria do autor.

No questionário também foi perguntado ao público se a aplicação teve bons resultados de previsão de acordo com cada celebridade, isso, visando a comparação de qual etnia o modelo tendia a obter melhores resultados de previsão. Os resultados da questão são apresentados na Figura 45. A avaliação feita na imagem gerada para o Will Smith, mostrou que os resultados não foram satisfatórios, e isso devido a imagem gerada realmente estar distante da aparência original do ator.

Em relação à celebridade Eunice Baía, foi quem obteve mais respostas positivas, o que vai em encontro aos resultados produzidos pelo modelo gerador, pois apresentaram resultados bastante precisos. As fotos originais da atriz também aparentavam que não estava sendo usado maquiagem ou que tenha passado por algum procedimento estético ou cirúrgico, desta forma o modelo

gerador seguiu as características identificadas para fazer a previsão, se aproximando da aparência da imagem original. Para as outras duas celebridades, Daniele Suzuki e Leonardo Dicaprio, também foi possível identificar que os resultados foram muito positivos, mas não tanto quanto os de Eunice Baía.

Figura 45 – Resultados para a qualidade da previsão de envelhecimento do modelo gerador por celebridade.



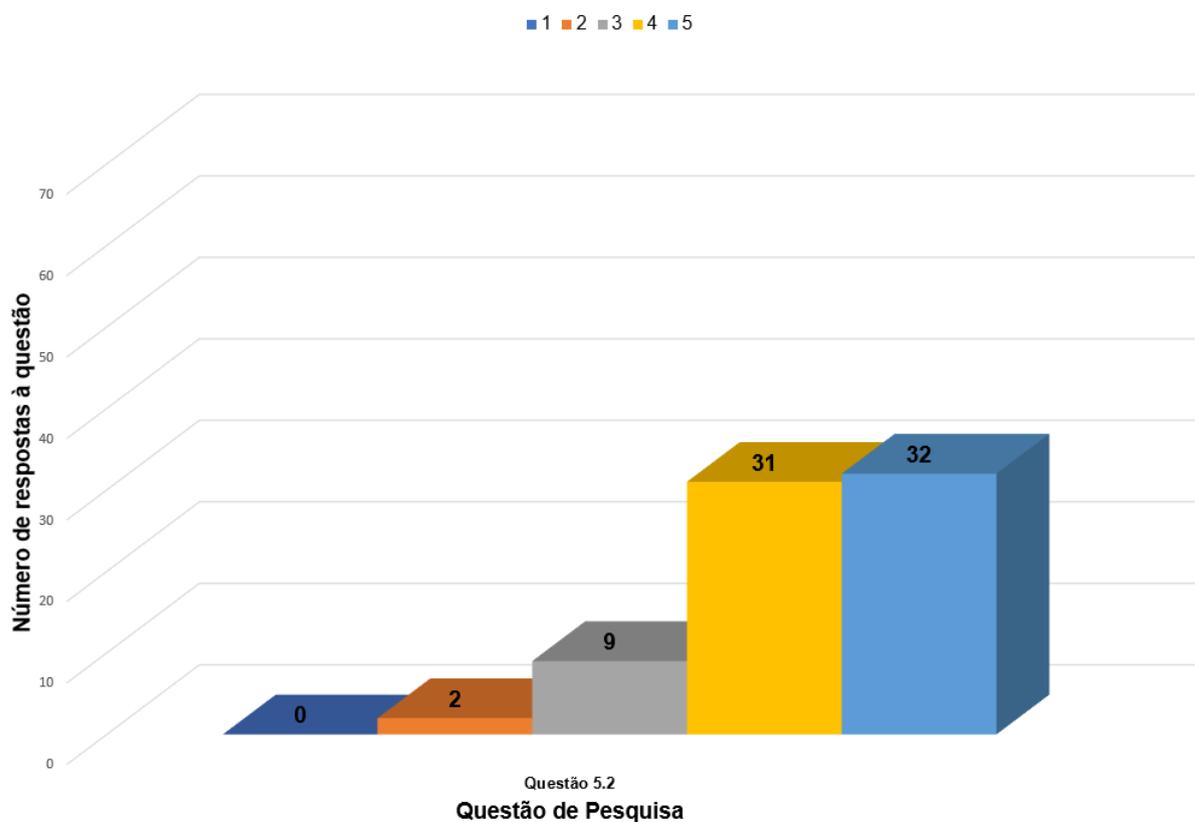
Fonte: Própria do autor.

A próxima etapa do questionário, se trata do grau de conhecimento/expertise de cada entrevistador, o que ajuda a ter uma noção de quantos entrevistados possuíam conhecimento para avaliar as imagens. Desta forma, foi possível perceber que mais de 65% dos entrevistados nunca trabalharam com edição de imagens ou utilizaram softwares para manipulação de imagens. Em contrapartida, a parcela restante disseram que já trabalharam ou utilizaram softwares para manipulação de imagens por cerca de mais de 1 ano, 3 anos, 4 anos ou mais de 5 anos. No entanto, grande parte desse público já utilizaram

ou utilizam redes sociais que contêm filtros de manipulação de imagem, como Instagram, Facebook, Twitter, Tiktok, Kwai e Snapchat. Dentre estes Apps, os mais utilizados pelos entrevistados são Instagram e Facebook.

Por fim, foi avaliado no questionário o nível de satisfação do público com relação a aplicação proposta neste trabalho. Quando foi perguntado se a aplicação cumpre o proposto, a grande maioria dos entrevistados, cerca de 92%, disseram que a aplicação cumpre o que promete. O nível de satisfação dos entrevistados é medido majoritariamente de forma quase que equiparada em satisfeito e muito satisfeito. O que pontua a aplicação proposta é uma ferramenta promissora para auxiliar na problemática abordada neste trabalho, podendo melhorar ainda mais com relação aos resultados produzidos. Os resultados de satisfação podem ser observados pela Figura 46.

Figura 46 – Resultados do nível de satisfação com a aplicação



Fonte: Própria do autor.

6 Considerações Finais e Trabalhos Futuros

Este trabalho apresentou uma abordagem e desenvolvimento de um software gerador de imagens faciais, usando StyleGAN com janelas de tempo, com a finalidade de auxiliar na busca por desaparecidos. O software proposto, consiste em a partir de uma imagem facial gerar uma nova imagem baseado em previsão temporal, ou seja, uma idade futura possível para uma dada pessoa desaparecida. Analisando trabalhos correlatos (ver Capítulo 3), foi identificado tipos de GANs e datasets que consolidaram a arquitetura do software proposto.

Na fase de avaliação do modelo, os resultados obtidos através das métricas FID, PPL e separabilidade linear, mostraram que o desempenho ao se trabalhar com as imagens no espaço latente W é superior em relação ao espaço Z , o que corresponde ao que foi descrito por Karras et al. (2019a). Além disso, também pode ser percebido que a idade da pessoa está diretamente ligada ao desempenho de previsão, ou seja, crianças tornam o sistema mais suscetível a erros de previsão, por não apresentarem traços tão bem definidos aparentes.

O questionário online aplicado também na fase de avaliação, mostrou que a aplicação obteve resultados significativos, como a comprovação da qualidade das imagens gerada pela aplicação e principalmente a aprovação do público para a aplicação e para os resultados preditivos gerados pelo modelo gerador. Porém, também é possível perceber pelos resultados do modelo e do questionário, que a aplicação possui muitos pontos ainda a serem melhorados, visto que a aplicação não é capaz de prever procedimentos estéticos, cirúrgicos ou maquiagens, por exemplo. Além disso, fatores de ambiente também podem afetar no resultado, como uma iluminação mais clara ou escura, que pode tornar a cor dos olhos, cabelos e tom da pele, mais claros ou escuros.

Os resultados obtidos com relação ao nível de precisão da previsão das imagens, foram satisfatórios, mostrando que a aplicação é capaz de gerar e

prever como um rosto humano será a partir da foto de entrada reconstruída. Sendo assim capaz de auxiliar no propósito da problemática deste trabalho. Porém, ainda sim existem muitos pontos a serem melhorados como a melhor preservação das características faciais e do gênero. Uma possível solução para que a preservação das características faciais e do gênero possa ser feita, é a adaptação da técnica de preservação de identidade, vista em (WANG et al., 2018), para o modelo da StyleGAN.

Visando melhorar a aplicação, para trabalhos futuros, pretende-se realizar um treinamento no modelo com uma maior variabilidade de etnias, acessórios faciais, condições de iluminação e gênero. Esta variabilidade se estende para fotos de crianças, desta forma visando melhorar os resultados da aplicação para esta faixa etária, uma vez que foi percebido durante a avaliação experimental que o sistema ao rejuvenescer um determinado rosto humano, tendia a errar e aproximar as características do indivíduo à de um asiático, mesmo a pessoa não sendo.

Outra pretensão visando melhorar os resultados é a adaptação da técnica de preservação de identidade, vista em (WANG et al., 2018), para o modelo de StyleGAN, afim de manter as características faciais do pessoa. Outra melhoria ao software proposto é a adição de um classificador para obter resultados com precisão como a acurácia das imagens geradas em relação às originais e também para a classificação das imagens geradas por idade da face. Por fim, a adição de um sistema Web interativo onde o próprio público poderia interagir com o sistema adicionando uma foto de entrada e obtendo o resultado de previsão das imagens geradas.

Referências

- AGARWAL, N.; BULLINS, B.; CHEN, X.; HAZAN, E.; SINGH, K.; ZHANG, C.; ZHANG, Y. **Efficient Full-Matrix Adaptive Regularization**. arXiv, 2020. Disponível em: <<https://arxiv.org/abs/1806.02958v2>>.
- AGGARWAL, C. C. **Neural Networks and Deep Learning: A Textbook**. 1st. ed. Cham: Springer Publishing Company, Incorporated, 2018. ISBN 3319944622.
- AMOS, B.; LUDWICZUK, B.; SATYANARAYANAN, M. et al. Openface: A general-purpose face recognition library with mobile applications. **CMU School of Computer Science**, v. 6, n. 2, 2016. Disponível em: <<http://reports-archive.adm.cs.cmu.edu/anon/anon/usr0/ftp/2016/CMU-CS-16-118.pdf>>.
- ANTIPOV, G.; BACCOUCHE, M.; BERRANI, S.-A.; DUGELAY, J.-L. Apparent age estimation from face images combining general and children-specialized deep learning models. **Proceedings of the IEEE conference on computer vision and pattern recognition workshops**, IEEE, p. 96–104, 2016. Disponível em: <https://openaccess.thecvf.com/content_cvpr_2016_workshops/w18/html/Antipov_Apparent_Age_Estimation_CVPR_2016_paper.html>.
- ANTIPOV, G.; BACCOUCHE, M.; DUGELAY, J.-L. Face aging with conditional generative adversarial networks. **2017 IEEE International Conference on Image Processing (ICIP)**, IEEE, p. 2089–2093, Sept. 2017. Disponível em: <<https://ieeexplore.ieee.org/document/8296650>>.
- AWAD, M.; KHANNA, R. Machine learning. In: _____. **Efficient Learning Machines: Theories, Concepts, and Applications for Engineers and System Designers**. Berkeley, CA: Apress, 2015. p. 1–18. ISBN 978-1-4302-5990-9. Disponível em: <https://doi.org/10.1007/978-1-4302-5990-9_1>.
- Cengil, E.; Çınar, A.; Özbay, E. Image classification with caffe deep learning framework. In: **2017 International Conference on Computer Science and Engineering (UBMK)**. IEEE, 2017. p. 440–444. Disponível em: <<https://ieeexplore.ieee.org/document/8093433>>.
- CHEN, B.-C.; CHEN, C.-S.; HSU, W. H. Face recognition and retrieval using cross-age reference coding with cross-age celebrity dataset. **IEEE Transactions on Multimedia**, IEEE, v. 17, n. 6, p. 804–815, 2015. Disponível em: <<https://ieeexplore.ieee.org/abstract/document/7080893>>.
- CHOLLET, F. **Deep Learning with Python**. 1st. ed. USA: Manning Publications Co., 2017. ISBN 1617294438.
- COELHO, N. Inteligência artificial e computação cognitiva em unidades de informação: conceitos e experiências. **Logeion: Filosofia da Informação**, v. 7, n. 1, p. 186–205, set. 2020. Disponível em: <<http://revista.ibict.br/fiinf/article/view/5260>>.

DING, X.; RAZIEI, Z.; LARSON, E. C.; OLINICK, E. V.; KRUEGER, P.; HAHLER, M. Swapped face detection using deep learning and subjective assessment. **CoRR**, abs/1909.04217, 2019. Disponível em: <<http://arxiv.org/abs/1909.04217>>.

GAUTHIER, J. Conditional generative adversarial nets for convolutional face generation. **Class Project for Stanford CS231N: Convolutional Neural Networks for Visual Recognition, Winter semester**, Winter semester, v. 2014, n. 5, p. 2, 2014. Disponível em: <<https://www.foldl.me/uploads/2015/conditional-gans-face-generation/paper.pdf>>.

GLOBALMISSINGKIDS. **Estatísticas de crianças desaparecidas**. International Centre for Missing & Exploited Children, 2020. [Online; Acessado em Abril de 2020]. Disponível em: <<https://globalmissingkids.org/awareness/missing-children-statistics/>>.

GOODFELLOW, I.; BENGIO, Y.; COURVILLE, A. **Deep Learning**. Cambridge, MA, USA: The MIT Press, 2016. ISBN 0262035618. Disponível em: <<http://www.deeplearningbook.org>>.

GOOGLE-COLAB. **Frequently Asked Questions**. Google Colaboratory, 2021. [Online; Acessado em Maio de 2021]. Disponível em: <<https://research.google.com/colaboratory/faq.html#gpu-availability>>.

HE, K.; ZHANG, X.; REN, S.; SUN, J. Deep residual learning for image recognition. In: **2016 IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2016, Las Vegas, NV, USA, June 27-30, 2016**. IEEE Computer Society, 2016. p. 770-778. Disponível em: <<https://doi.org/10.1109/CVPR.2016.90>>.

IBM. **20 Popular Machine Learning Metrics. Part 1: Classification & Regression Evaluation Metrics**. IBM, 2021. [Online; Acessado em Março de 2021]. Disponível em: <<https://www.ibm.com/topics/computer-vision>>.

JIA, Y.; SHELHAMER, E.; DONAHUE, J.; KARAYEV, S.; LONG, J.; GIRSHICK, R.; GUADARRAMA, S.; DARRELL, T. Caffe: Convolutional architecture for fast feature embedding. **CoRR**, abs/1408.5093, 2014. Disponível em: <<http://arxiv.org/abs/1408.5093>>.

JIN, Y.; ZHANG, J.; LI, M.; TIAN, Y.; ZHU, H.; FANG, Z. Towards the automatic anime characters creation with generative adversarial networks. **CoRR**, abs/1708.05509, 2017. Disponível em: <<http://arxiv.org/abs/1708.05509>>.

JOSHI, N. **5 applications of generative adversarial networks**. Allerin, 2020. [Online; Acessado em Março de 2021]. Disponível em: <<https://www.allerin.com/blog/5-applications-of-generative-adversarial-networks>>.

KAPOOR, A.; GUILI, A.; PAL, S. **Deep Learning with TensorFlow 2 and Keras: Regression, ConvNets, GANs, RNNs, NLP, and more with TensorFlow 2 and the Keras API, 2nd Edition**. Birmingham, UK: Packt Publishing, 2019. v. 2. ISBN 9781838823412.

KARRAS, T.; AILA, T.; LAINE, S.; LEHTINEN, J. Progressive growing of gans for improved quality, stability, and variation. **CoRR**, abs/1710.10196, 2018. Disponível em: <<http://arxiv.org/abs/1710.10196>>.

KARRAS, T.; LAINE, S.; AILA, T. A style-based generator architecture for generative adversarial networks. In: **IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2019, Long Beach, CA, USA, June 16-20, 2019**. Computer Vision Foundation / IEEE, 2019. p. 4401–4410. Disponível em: <http://openaccess.thecvf.com/content_CVPR_2019/html/Karras_A_Style-Based_Generator_Architecture_for_Generative_Adversarial_Networks_CVPR_2019_paper.html>.

KARRAS, T.; LAINE, S.; AILA, T. **A Style-Based Generator Architecture for Generative Adversarial Networks**. 2019. Disponível em <https://www.slideserve.com/vfenton/a-style-based-generator-architecture-for-generative-adversarial-networks-powerpoint-ppt-presentation> e também em https://aisc.ai/science/static/slides/20190124_DiegoCantor.pdf. [Online; Acessado em Setembro de 2021].

KERAS. **R Interface to Keras**. Keras, 2021. [Online; Acessado em Março de 2021]. Disponível em: <<https://keras.rstudio.com/>>.

KOK, J. N.; BOERS, E. J. W.; KOSTERS, W. A.; PUTTEN, P. van der. **Artificial Intelligence: Definition, Trends, Techniques and Cases**. Eolss Publishers, 2009. [Online; Acessado em Março de 2021]. Disponível em: <<https://www.eolss.net/Sample-Chapters/C15/E6-44.pdf>>.

KUMAR, A.; SALAU, A.; GUPTA, S.; ARORA, S. A survey of machine learning methods for iot and their future applications. **Amity Journal of Computational Sciences (AJCS)**, v. 2, n. 2, p. 1–5, 01 2018. Disponível em: <https://www.researchgate.net/publication/330761913_A_Survey_of_Machine_Learning_Methods_for_IoT_and_their_Future_Applications>.

LI, P.; HU, Y.; LI, Q.; HE, R.; SUN, Z. Global and local consistent age generative adversarial networks. In: **2018 24th International Conference on Pattern Recognition (ICPR)**. IEEE, 2018. p. 1073–1078. Disponível em: <<https://ieeexplore.ieee.org/abstract/document/8545119>>.

LIU, S.; SUN, Y.; ZHU, D.; BAO, R.; WANG, W.; SHU, X.; YAN, S. Face aging with contextual generative adversarial nets. In: **MM '17: Proceedings of the 25th ACM international conference on Multimedia**. New York, NY, USA: Association for Computing Machinery, 2017. (MM '17), p. 82–90. ISBN 9781450349062. Disponível em: <<https://doi.org/10.1145/3123266.3123431>>.

LIU, Y.; LI, Q.; SUN, Z. Attribute-aware face aging with wavelet-based generative adversarial networks. In: **Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)**. IEEE, 2019. p. 11869–11878. Disponível em: <<https://ieeexplore.ieee.org/document/8954076>>.

LIU, Z.; LUO, P.; WANG, X.; TANG, X. Deep learning face attributes in the wild. In: **2015 IEEE International Conference on Computer Vision, ICCV 2015, Santiago, Chile, December 7-13, 2015**. IEEE Computer Society, 2015. p. 3730–3738. Disponível em: <<https://doi.org/10.1109/ICCV.2015.425>>.

- LUGER, G. F. **Inteligência Artificial - 4.ED.: Estruturas e estratégias para a solução de problemas complexos**. Brasil: Editora Bookman, 2004. v. 4^a ed. 774 p. ISBN 9788577801206.
- MEGVII-INC. **Face++ research toolkit**. Megvii Inc., 2021. [Online; Acessado em Março de 2021]. Disponível em: <<https://www.faceplusplus.com>>.
- MIRZA, M.; OSINDERO, S. Conditional generative adversarial nets. **CoRR**, abs/1411.1784, 2014. Disponível em: <<http://arxiv.org/abs/1411.1784>>.
- NGUYEN, T. T.; NGUYEN, C. M.; NGUYEN, D. T.; NGUYEN, D. T.; NAHAVANDI, S. Deep learning for deepfakes creation and detection: A survey. **CoRR**, abs/1909.11573, Jul 2020. Disponível em: <<http://arxiv.org/abs/1909.11573>>.
- NIKITKO, D. **StyleGAN — Encoder for Official TensorFlow Implementation**. 2019. <<https://github.com/Puzer/stylegan-encoder>>. [Online; Acessado em Julho de 2021].
- PATTERSON, J.; GIBSON, A. **Deep Learning: A Practitioner's Approach**. 1st. ed. Sebastopol, CA: O'Reilly Media, Inc., 2017. ISBN 1491914254.
- RICANEK, K.; TESAFAYE, T. Morph: A longitudinal image database of normal adult age-progression. In: **7th International Conference on Automatic Face and Gesture Recognition (FG06)**. IEEE, 2006. p. 341–345. Disponível em: <<https://ieeexplore.ieee.org/document/1613043>>.
- RICH, E.; KNIGHT, K. **Inteligência Artificial**. São Paulo: McGraw-Hill, 1994. v. 2^a ed. 722 p.
- 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)**. IEEE, 2015. p. 252–257. Disponível em: <<https://ieeexplore.ieee.org/document/7406390>>.
- SAMUEL, A. L. Some studies in machine learning using the game of checkers. **IBM Journal of research and development**, IBM, v. 3, n. 3, p. 210–229, 1959. Disponível em: <<https://ieeexplore.ieee.org/abstract/document/5392560/>>.
- SHEN, Y.; GU, J.; TANG, X.; ZHOU, B. Interpreting the latent space of gans for semantic face editing. In: **2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition, CVPR 2020, Seattle, WA, USA, June 13-19, 2020**. IEEE, 2020. p. 9240–9249. Disponível em: <<https://doi.org/10.1109/CVPR42600.2020.00926>>.
- VASILEV, I.; SLATER, D.; SPACAGNA, G.; ROELANTS, P.; ZOCCA, V. **Python Deep Learning: Exploring deep learning techniques and neural network architectures with PyTorch, Keras, and TensorFlow, 2nd Edition**. Birmingham, UK: Packt Publishing, 2019. v. 2^a ed. ISBN 9781789349702.
- VIEIRA, A. **Desaparecimento de crianças: Conselho de Medicina faz alerta na Comissão de Direitos Humanos**. Agência Senado, 2018. [Online; Acessado em Abril de 2020]. Disponível

em: <<https://www12.senado.leg.br/noticias/materias/2018/04/06/desaparecimento-de-criancas-conselho-de-medicina-faz-alerta-na-comissao-de-direitos-humanos>>.

WANG, Z.; SHE, Q.; WARD, T. E. Generative adversarial networks in computer vision: A survey and taxonomy. **CoRR**, abs/1906.01529, 2020. Disponível em: <<http://arxiv.org/abs/1906.01529>>.

WANG, Z.; TANG, X.; LUO, W.; GAO, S. Face aging with identity-preserved conditional generative adversarial networks. In: **Proceedings of the IEEE conference on computer vision and pattern recognition**. [S.l.: s.n.], 2018. p. 7939–7947.

WU, J.; ZHANG, C.; XUE, T.; FREEMAN, W. T.; TENENBAUM, J. B. Learning a probabilistic latent space of object shapes via 3d generative-adversarial modeling. **CoRR**, abs/1610.07584, Jan 2017. Disponível em: <<http://arxiv.org/abs/1610.07584>>.

WU, X.; XU, K.; HALL, P. A survey of image synthesis and editing with generative adversarial networks. **Tsinghua Science and Technology**, Tsinghu University Press, v. 22, n. 6, p. 660–674, 2017. Disponível em: <<https://ieeexplore.ieee.org/document/8195348>>.

YAMASHITA, R.; NISHIO, M.; DO, R.; TOGASHI, K. Convolutional neural networks: an overview and application in radiology. **Insights into Imaging**, v. 9, n. 4, p. 611–629, 06 2018. Disponível em: <https://www.researchgate.net/publication/325932686_Convolutional_neural_networks_an_overview_and_application_in_radiology>.

YANG, H.; HUANG, D.; WANG, Y.; JAIN, A. K. Learning face age progression: A pyramid architecture of gans. **Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)**, IEEE, p. 31–39, June 2018.

YI, R.; LIU, Y.-J.; LAI, Y.-K.; ROSIN, P. L. Apdrawinggan: Generating artistic portrait drawings from face photos with hierarchical gans. **2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)**, IEEE, p. 10735–10744, June 2019. Disponível em: <<https://ieeexplore.ieee.org/document/8953680>>.

ZAKHAROV, E.; SHYSHEYA, A.; BURKOV, E.; LEMPITSKY, V. Few-shot adversarial learning of realistic neural talking head models. **2019 IEEE/CVF International Conference on Computer Vision (ICCV)**, p. 9458–9467, 2019. Disponível em: <<https://ieeexplore.ieee.org/document/9009591>>.

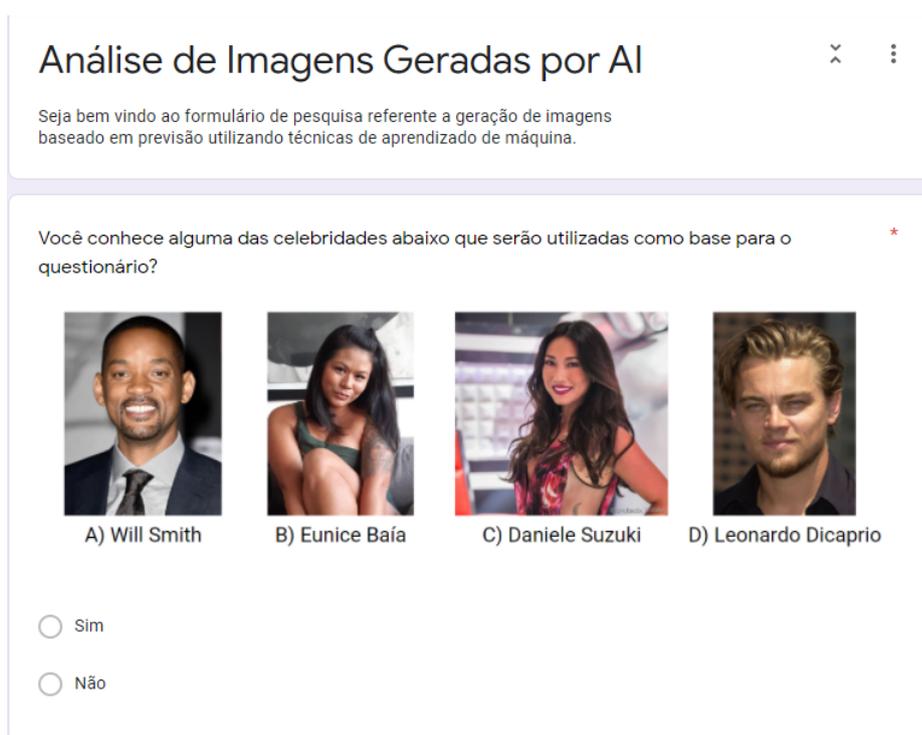
APÊNDICE A – Questionário

Este apêndice apresenta a estrutura do questionário *online* que foi aplicado ao público, as questões de pesquisa que o compõem, bem como, os resultados coletados da avaliação.

A.1 Estrutura do questionário e questões de pesquisa

Nas figuras seguintes pode ser observado a estrutura do questionário e a composição de questões do questionário *online* aplicado.

Figura 47 – Início do questionário - parte 1



The image shows a screenshot of a web-based questionnaire. At the top, the title is "Análise de Imagens Geradas por AI" with a close button (X) and a menu icon (three dots). Below the title, a welcome message reads: "Seja bem vindo ao formulário de pesquisa referente a geração de imagens baseado em previsão utilizando técnicas de aprendizado de máquina." The main question is: "Você conhece alguma das celebridades abaixo que serão utilizadas como base para o questionário?" followed by a red asterisk. Below the question are four portrait photos of celebrities: A) Will Smith, B) Eunice Baía, C) Daniele Suzuki, and D) Leonardo Dicaprio. At the bottom, there are two radio button options: "Sim" and "Não".

Fonte: Própria do autor.

Figura 48 – Início do questionário - parte 2

Se sim, qual ou quais delas conhece? *

Will Smith

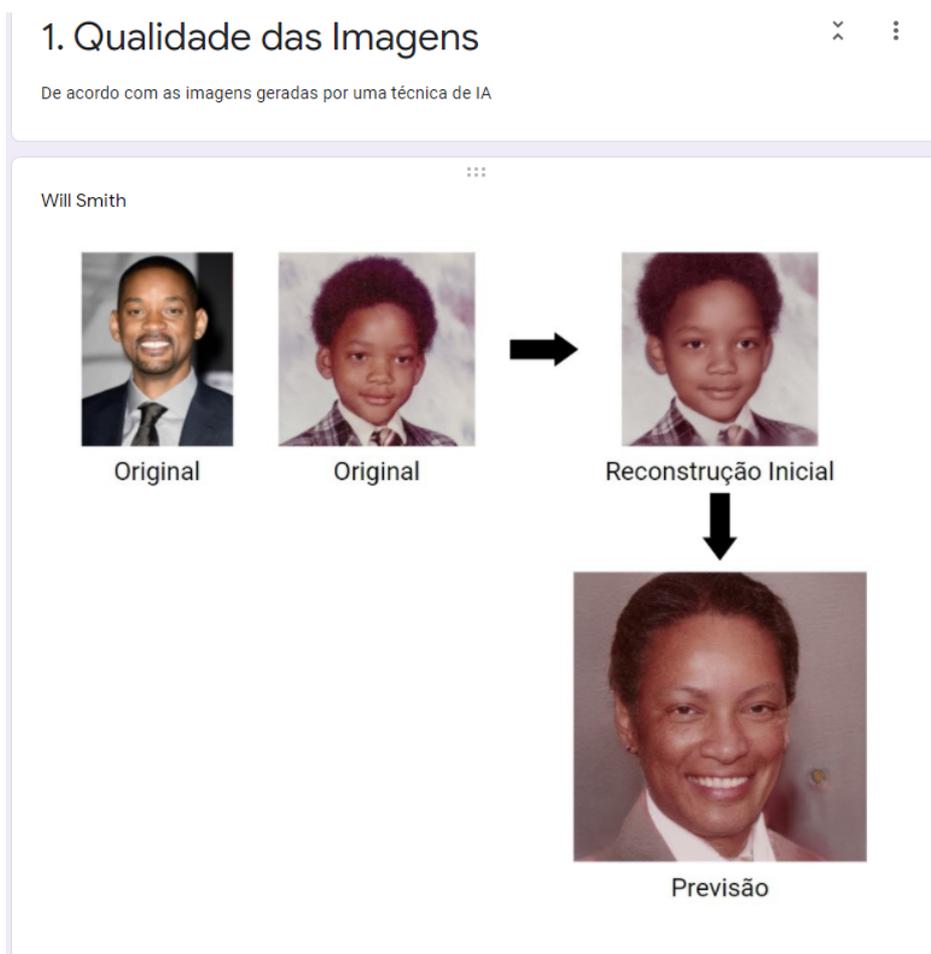
Eunice Baía

Daniele Suzuki

Leonardo Dicaprio

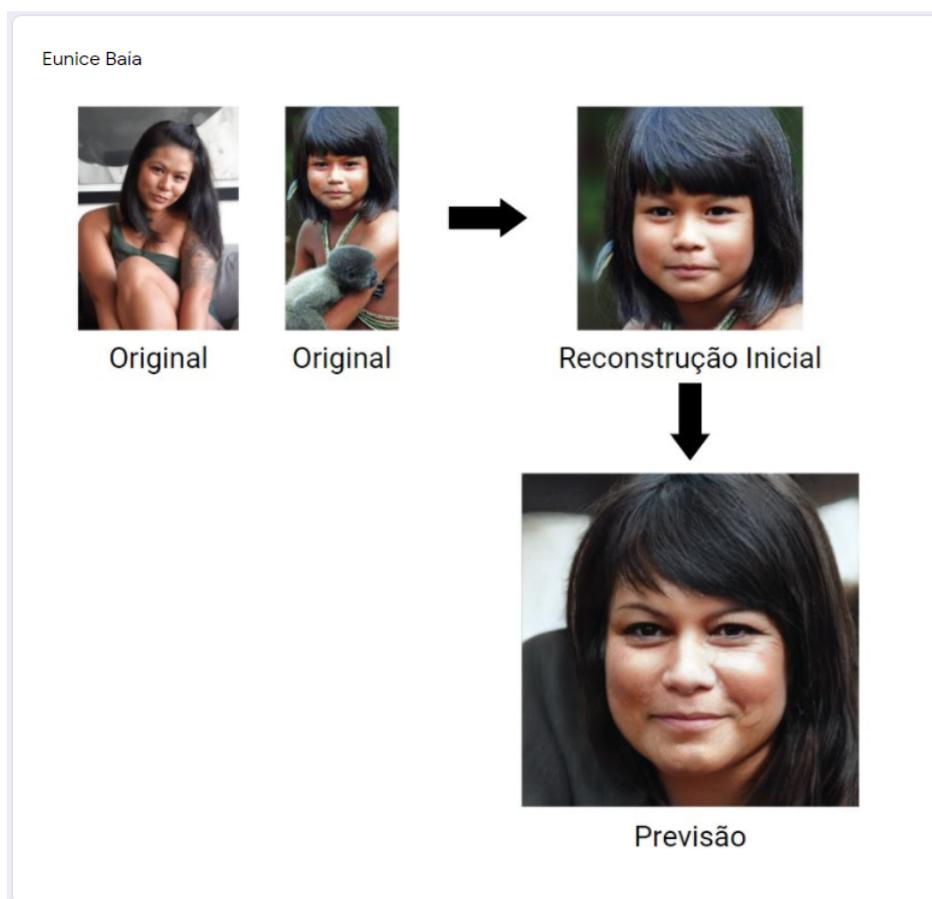
Fonte: Própria do autor.

Figura 49 – Qualidade das imagens - parte 1



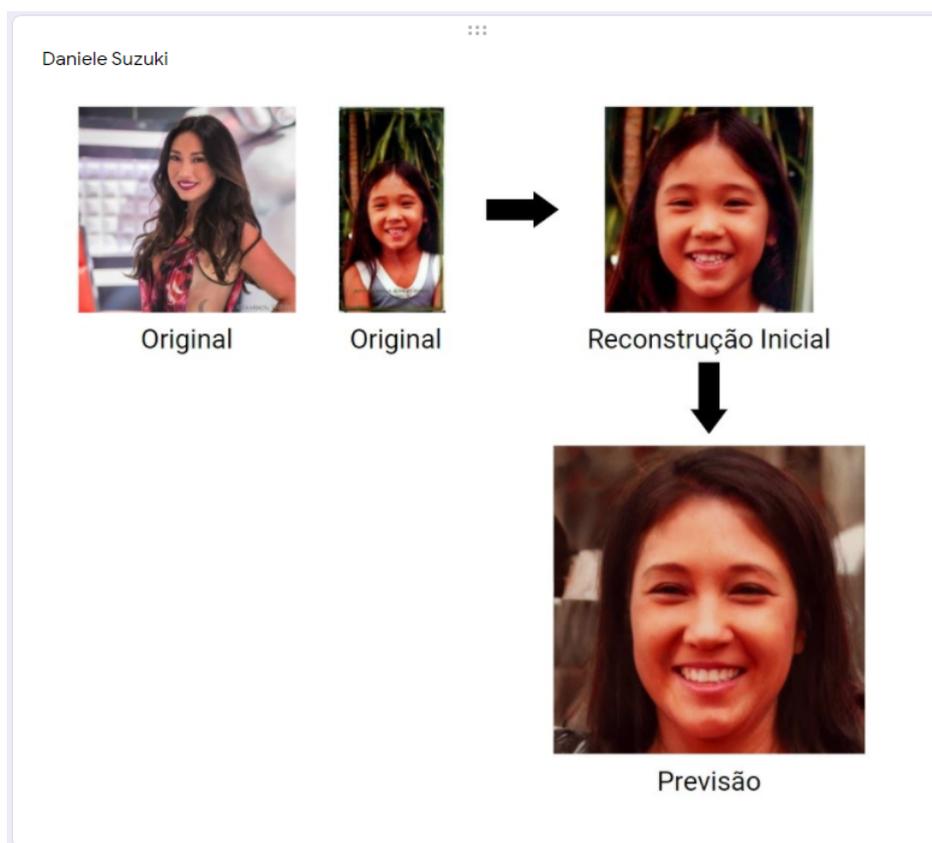
Fonte: Própria do autor.

Figura 50 – Qualidade das imagens - parte 2



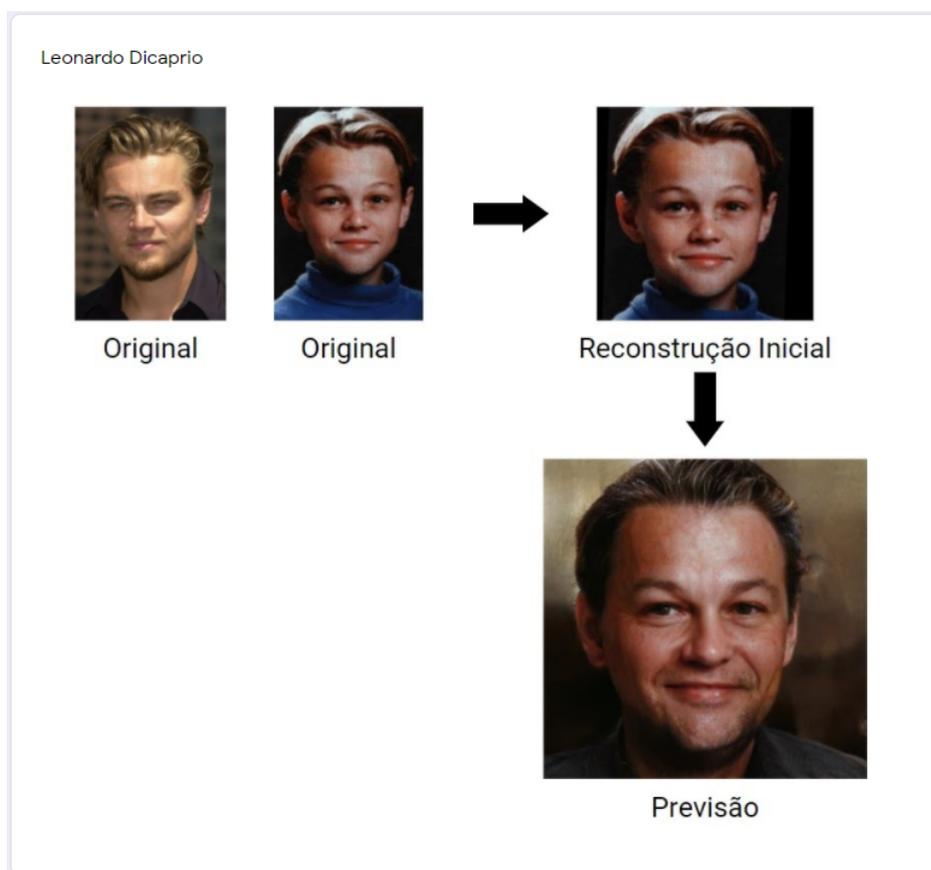
Fonte: Própria do autor.

Figura 51 – Qualidade das imagens - parte 3



Fonte: Própria do autor.

Figura 52 – Qualidade das imagens - parte 4



Fonte: Própria do autor.

Figura 53 – Qualidade das imagens - parte 5

1.1 Como você classifica o nível de nitidez das imagens geradas? *						
	1	2	3	4	5	
Muito Ruim	<input type="radio"/>	Muito Bom				

1.2 Como você classifica o nível de saturação (cores vivas) das imagens geradas? *						
	1	2	3	4	5	
Muito Ruim	<input type="radio"/>	Muito Bom				

1.3 Quanto você considera que os rostos gerados nas fotos se parecem com os artistas? *						
	1	2	3	4	5	
Nada	<input type="radio"/>	Totalmente Parecido				

Fonte: Própria do autor.

Figura 54 – Características Faciais - Will Smith - parte 1

2. Características Faciais ✕ ⋮

De acordo com as fotos geradas pela aplicação e a celebridade a baixo, responda as questões que se seguem.

Will Smith



Original



Original

➔



Reconstrução Inicial

↓



Previsão

Fonte: Própria do autor.

Figura 55 – Características Faciais - Will Smith - parte 2

Com relação a face como um todo						
Descrição (opcional)						
2.1. Como você avalia o contorno (normal ou anormal) do rosto gerado nas imagens geradas? *						
	1	2	3	4	5	
Muito Ruim	<input type="radio"/>	Muito Bom				
Com relação aos olhos						
Descrição (opcional)						
2.2. Como você avalia a qualidade dos detalhes (expressão ou traços) gerados dos olhos? *						
	1	2	3	4	5	
Muito Ruim	<input type="radio"/>	Muito Bom				
2.3. Como você avalia o nível de semelhança da cor dos olhos? *						
	1	2	3	4	5	
Muito Ruim	<input type="radio"/>	Muito Bom				

Fonte: Própria do autor.

Figura 56 – Características Faciais - Will Smith - parte 3

Com relação ao nariz						
Descrição (opcional)						
2.4. Como você avalia os detalhes (largura e formato) gerados para o nariz? *						
	1	2	3	4	5	
Muito Ruim	<input type="radio"/>	Muito Bom				
Com relação as orelhas						
Descrição (opcional)						
2.5. Como você avalia os detalhes (comprimento, curvatura) gerados para as orelhas? *						
	1	2	3	4	5	
Muito Ruim	<input type="radio"/>	Muito Bom				

Fonte: Própria do autor.

Figura 57 – Características Faciais - Will Smith - parte 4

Com relação a boca						
Descrição (opcional)						
2.6. Como você avalia os detalhes (traços) gerados para a boca? *						
	1	2	3	4	5	
Muito Ruim	<input type="radio"/>	Muito Bom				
Com relação ao cabelo						
Descrição (opcional)						
2.7. Como você avalia a cor do cabelo? *						
	1	2	3	4	5	
Muito Ruim	<input type="radio"/>	Muito Bom				
2.8. Como você avalia o comprimento do cabelo? *						
	1	2	3	4	5	
Muito Ruim	<input type="radio"/>	Muito Bom				

Fonte: Própria do autor.

Figura 58 – Características Faciais - Will Smith - parte 5

2.9. Como você avalia o volume do cabelo? *						
	1	2	3	4	5	
Muito Ruim	<input type="radio"/>	Muito Bom				

2.10. Como você avalia o corte do cabelo? *						
	1	2	3	4	5	
Muito Ruim	<input type="radio"/>	Muito Bom				

Fonte: Própria do autor.

Figura 59 – Características Faciais - Eunice Baía - parte 1

2. Características Faciais ✕ ⋮

De acordo com as fotos geradas pela aplicação e a celebridade a baixo, responda as questões que se seguem.

Eunice Baía



Original



Original

➔



Reconstrução Inicial

↓



Previsão

Fonte: Própria do autor.

Figura 60 – Características Faciais - Eunice Baía - parte 2

Com relação a face como um todo						
Descrição (opcional)						
2.1. Como você avalia o contorno (normal ou anormal) do rosto gerado nas imagens geradas? *						
	1	2	3	4	5	
Muito Ruim	<input type="radio"/>	Muito Bom				
Com relação aos olhos						
Descrição (opcional)						
2.2. Como você avalia a qualidade dos detalhes (expressão ou traços) gerados dos olhos? *						
	1	2	3	4	5	
Muito Ruim	<input type="radio"/>	Muito Bom				
2.3. Como você avalia o nível de semelhança da cor dos olhos? *						
	1	2	3	4	5	
Muito Ruim	<input type="radio"/>	Muito Bom				

Fonte: Própria do autor.

Figura 61 – Características Faciais - Eunice Baía - parte 3

Com relação ao nariz						
Descrição (opcional)						
2.4. Como você avalia os detalhes (largura e formato) gerados para o nariz? *						
	1	2	3	4	5	
Muito Ruim	<input type="radio"/>	Muito Bom				

Com relação a boca						
Descrição (opcional)						
2.5. Como você avalia os detalhes (traços) gerados para a boca? *						
	1	2	3	4	5	
Muito Ruim	<input type="radio"/>	Muito Bom				

Fonte: Própria do autor.

Figura 62 – Características Faciais - Eunice Baía - parte 4

Com relação ao cabelo						
Descrição (opcional)						
2.6. Como você avalia a cor do cabelo? *						
	1	2	3	4	5	
Muito Ruim	<input type="radio"/>	Muito Bom				

2.7. Como você avalia o comprimento do cabelo? *						
	1	2	3	4	5	
Muito Ruim	<input type="radio"/>	Muito Bom				

Fonte: Própria do autor.

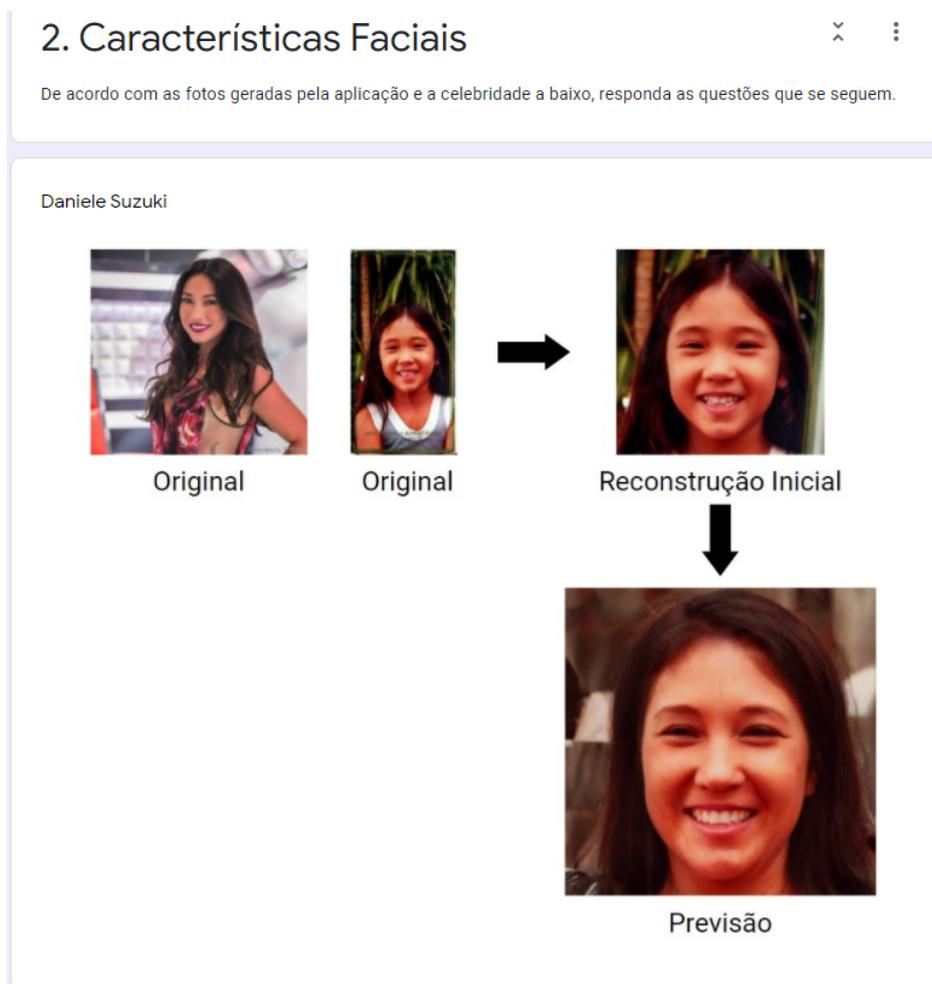
Figura 63 – Características Faciais - Eunice Baía - parte 5

2.8. Como você avalia o volume do cabelo? *						
	1	2	3	4	5	
Muito Ruim	<input type="radio"/>	Muito Bom				

2.9. Como você avalia o corte do cabelo? *						
	1	2	3	4	5	
Muito Ruim	<input type="radio"/>	Muito Bom				

Fonte: Própria do autor.

Figura 64 – Características Faciais - Daniele Suzuki - parte 1



Fonte: Própria do autor.

Figura 65 – Características Faciais - Daniele Suzuki - parte 2

Com relação a face como um todo						
Descrição (opcional)						
2.1. Como você avalia o contorno (normal ou anormal) do rosto gerado nas imagens geradas? *						
	1	2	3	4	5	
Muito Ruim	<input type="radio"/>	Muito Bom				
Com relação aos olhos						
Descrição (opcional)						
2.2. Como você avalia a qualidade dos detalhes (expressão ou traços) gerados dos olhos? *						
	1	2	3	4	5	
Muito Ruim	<input type="radio"/>	Muito Bom				
2.3. Como você avalia o nível de semelhança da cor dos olhos? *						
	1	2	3	4	5	
Muito Ruim	<input type="radio"/>	Muito Bom				

Fonte: Própria do autor.

Figura 66 – Características Faciais - Daniele Suzuki - parte 3

Com relação ao nariz

Descrição (opcional)

2.4. Como você avalia os detalhes (largura e formato) gerados para o nariz? *

1 2 3 4 5

Muito Ruim Muito Bom

Com relação a boca

Descrição (opcional)

2.5. Como você avalia os detalhes (traços) gerados para a boca? *

1 2 3 4 5

Muito Ruim Muito Bom

Fonte: Própria do autor.

Figura 67 – Características Faciais - Daniele Suzuki - parte 4

Com relação ao cabelo

Descrição (opcional)

2.6. Como você avalia a cor do cabelo? *

1 2 3 4 5

Muito Ruim Muito Bom

2.7. Como você avalia o comprimento do cabelo? *

1 2 3 4 5

Muito Ruim Muito Bom

Fonte: Própria do autor.

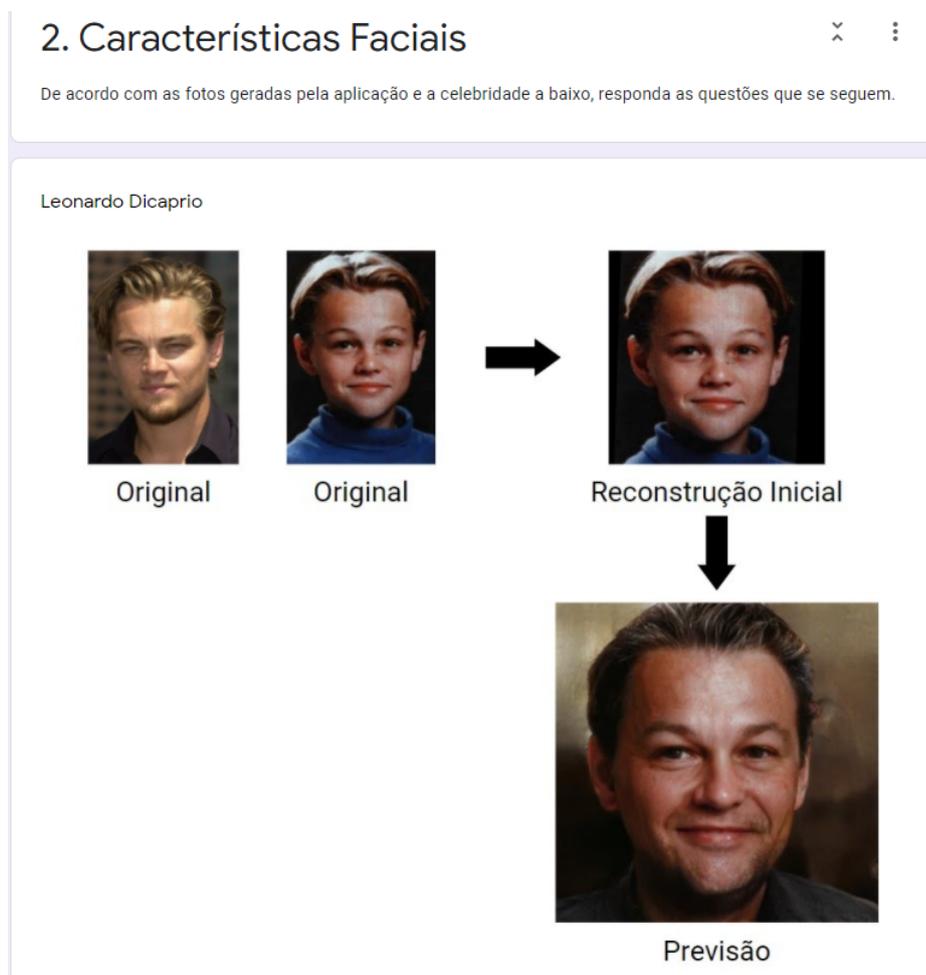
Figura 68 – Características Faciais - Daniele Suzuki - parte 5

2.8. Como você avalia o volume do cabelo? *						
	1	2	3	4	5	
Muito Ruim	<input type="radio"/>	Muito Bom				

2.9. Como você avalia o corte do cabelo? *						
	1	2	3	4	5	
Muito Ruim	<input type="radio"/>	Muito Bom				

Fonte: Própria do autor.

Figura 69 – Características Faciais - Leonardo Dicaprio - parte 1



Fonte: Própria do autor.

Figura 70 – Características Faciais - Leonardo Dicaprio - parte 2

Com relação a face como um todo						
Descrição (opcional)						
2.1. Como você avalia o contorno (normal ou anormal) do rosto gerado nas imagens geradas? *						
	1	2	3	4	5	
Muito Ruim	<input type="radio"/>	Muito Bom				
Com relação aos olhos						
Descrição (opcional)						
2.2. Como você avalia a qualidade dos detalhes (expressão ou traços) gerados dos olhos? *						
	1	2	3	4	5	
Muito Ruim	<input type="radio"/>	Muito Bom				
2.3. Como você avalia o nível de semelhança da cor dos olhos? *						
	1	2	3	4	5	
Muito Ruim	<input type="radio"/>	Muito Bom				

Fonte: Própria do autor.

Figura 71 – Características Faciais - Leonardo Dicaprio - parte 3

Com relação ao nariz						
Descrição (opcional)						
2.4. Como você avalia os detalhes (largura e formato) gerados para o nariz? *						
	1	2	3	4	5	
Muito Ruim	<input type="radio"/>	Muito Bom				
Com relação as orelhas						
Descrição (opcional)						
2.5. Como você avalia os detalhes (comprimento, curvatura) gerados para as orelhas? *						
	1	2	3	4	5	
Muito Ruim	<input type="radio"/>	Muito Bom				

Fonte: Própria do autor.

Figura 72 – Características Faciais - Leonardo Dicaprio - parte 4

Com relação a boca

Descrição (opcional)

2.6. Como você avalia os detalhes (traços) gerados para a boca? *

	1	2	3	4	5	
Muito Ruim	<input type="radio"/>	Muito Bom				

Com relação ao cabelo

Descrição (opcional)

2.7. Como você avalia a cor do cabelo? *

	1	2	3	4	5	
Muito Ruim	<input type="radio"/>	Muito Bom				

2.8. Como você avalia o comprimento do cabelo? *

	1	2	3	4	5	
Muito Ruim	<input type="radio"/>	Muito Bom				

Fonte: Própria do autor.

Figura 73 – Características Faciais - Leonardo Dicaprio - parte 5

2.9. Como você avalia o volume do cabelo? *

	1	2	3	4	5	
Muito Ruim	<input type="radio"/>	Muito Bom				

2.10. Como você avalia o corte do cabelo? *

	1	2	3	4	5	
Muito Ruim	<input type="radio"/>	Muito Bom				

Fonte: Própria do autor.

Figura 74 – Qualidade do envelhecimento - parte 1

3. Qualidade do Envelhecimento

Descrição (opcional)

Imagens geradas pela aplicação



A) Will Smith B) Eunice Baía C) Daniele Suzuki D) Leonardo DiCaprio

3.1. De acordo com as imagens, geradas pela aplicação, em qual faixa etária você às classifica? *

1+ Anos

20+ Anos

30+ Anos

40+ Anos

50+ Anos

Fonte: Própria do autor.

Figura 75 – Qualidade do envelhecimento - parte 2

3.2. Quanto você considera que as imagens geradas obtiveram uma boa previsão de envelhecimento? *						
	1	2	3	4	5	
Muito Ruim	<input type="radio"/>	Muito Bom				
Para cada uma das celebridades						
Descrição (opcional)						
3.3. De acordo com a celebridade "A" na figura, você considera que a aplicação teve bons resultados para a previsão ? *						
	1	2	3	4	5	
Ruim	<input type="radio"/>	Exelente				

Fonte: Própria do autor.

Figura 76 – Qualidade do envelhecimento - parte 3

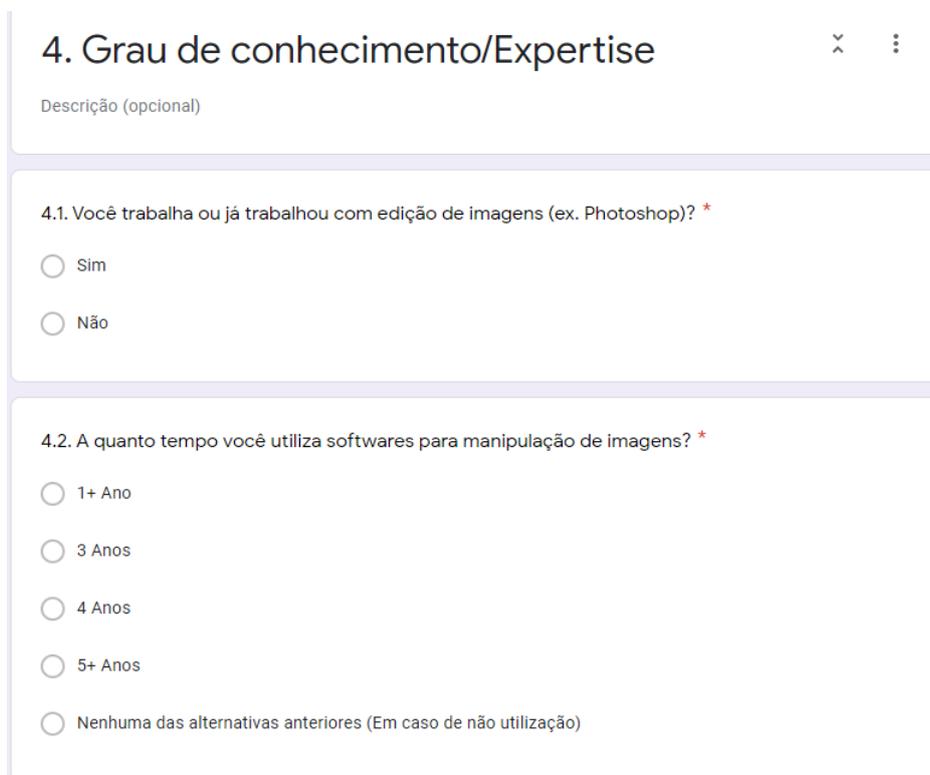
3.4. De acordo com a celebridade "B)" na figura, você considera que a aplicação teve bons resultados para a previsão ? *						
	1	2	3	4	5	
Ruim	<input type="radio"/>	Exelente				

3.5. De acordo com a celebridade "C)" na figura, você considera que a aplicação teve bons resultados para a previsão ? *						
	1	2	3	4	5	
Ruim	<input type="radio"/>	Exelente				

3.6. De acordo com a celebridade "D)" na figura, você considera que a aplicação teve bons resultados para a previsão ? *						
	1	2	3	4	5	
Ruim	<input type="radio"/>	Exelente				

Fonte: Própria do autor.

Figura 77 – Grau de conhecimento/Expertise - parte 1



4. Grau de conhecimento/Expertise

Descrição (opcional)

4.1. Você trabalha ou já trabalhou com edição de imagens (ex. Photoshop)? *

Sim

Não

4.2. A quanto tempo você utiliza softwares para manipulação de imagens? *

1+ Ano

3 Anos

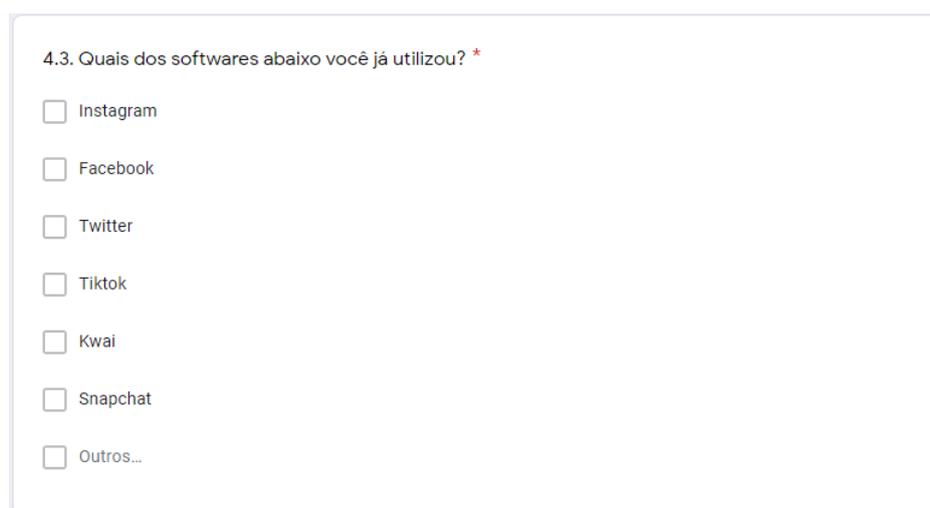
4 Anos

5+ Anos

Nenhuma das alternativas anteriores (Em caso de não utilização)

Fonte: Própria do autor.

Figura 78 – Grau de conhecimento/Expertise - parte 2



4.3. Quais dos softwares abaixo você já utilizou? *

Instagram

Facebook

Twitter

Tiktok

Kwai

Snapchat

Outros...

Fonte: Própria do autor.

Figura 79 – Satisfação

5. Satisfação

Descrição (opcional)

5.1. Na sua opinião, a aplicação cumpriu/cumpru o que promete? *

Sim

Não

5.2. O quão satisfeito você está com relação aos resultados produzidos pela aplicação? *

1 2 3 4 5

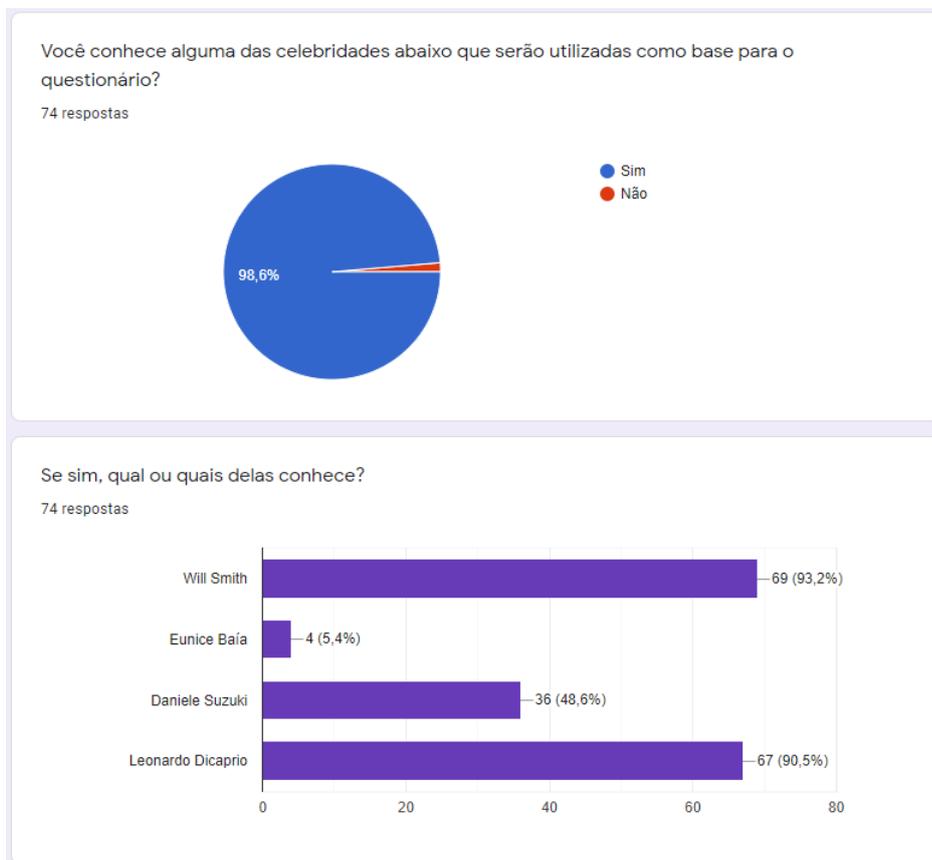
Nenhum pouco satisfeito Muito satisfeito

Fonte: Própria do autor.

A.2 Resultados coletados da avaliação

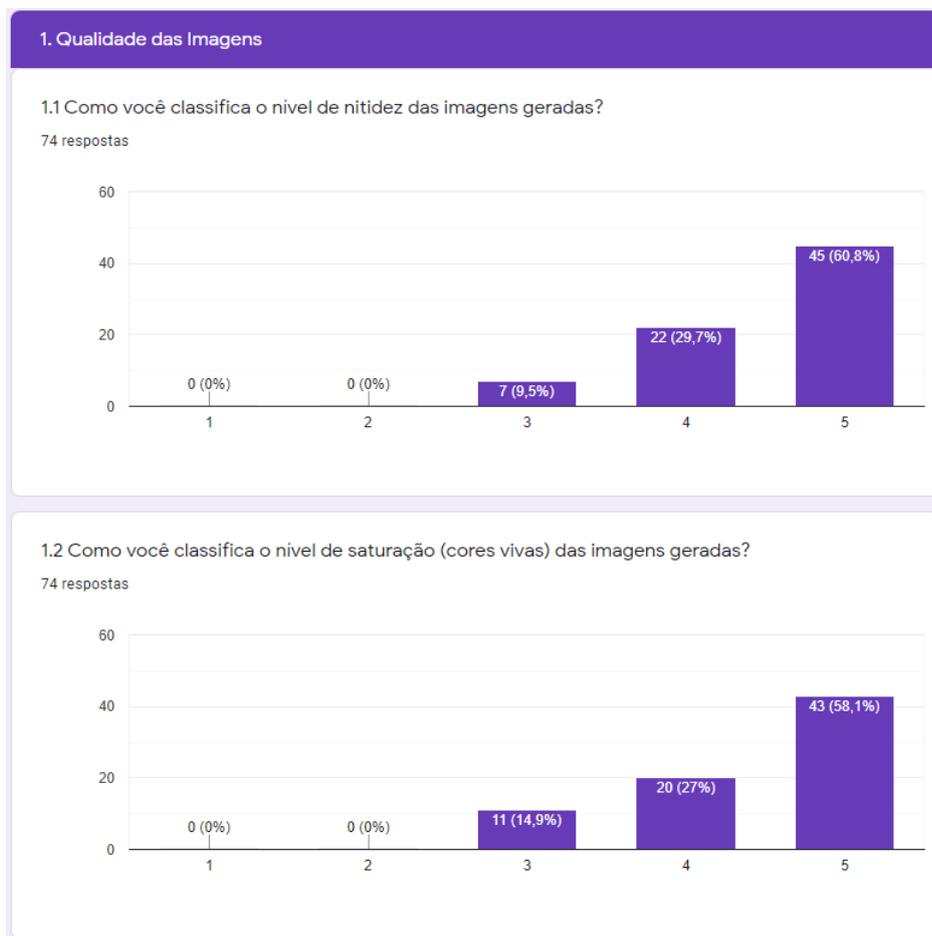
Nas figuras seguintes pode ser observado os resultados coletados da avaliação do questionário online aplicado.

Figura 80 – Resultados - Início do questionário



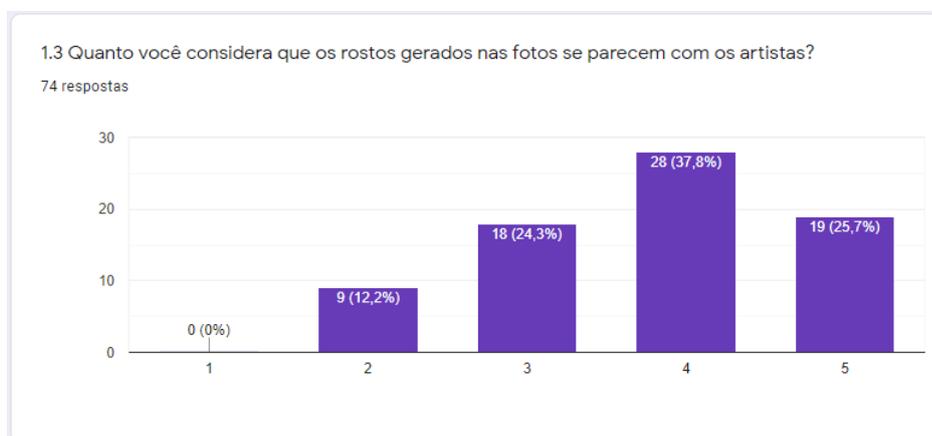
Fonte: Própria do autor.

Figura 81 – Resultados - Qualidade das imagens - parte 1



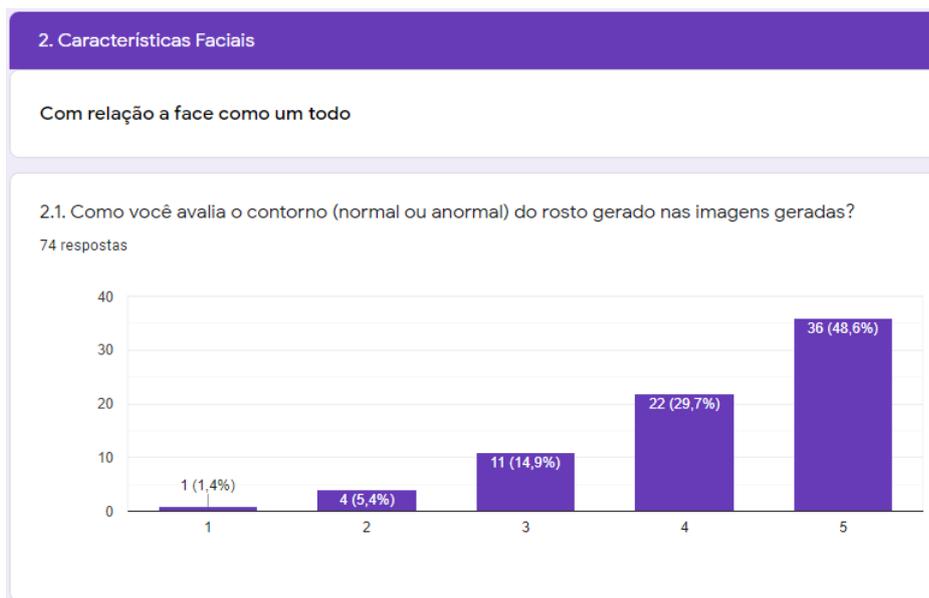
Fonte: Própria do autor.

Figura 82 – Resultados - Qualidade das imagens - parte 2



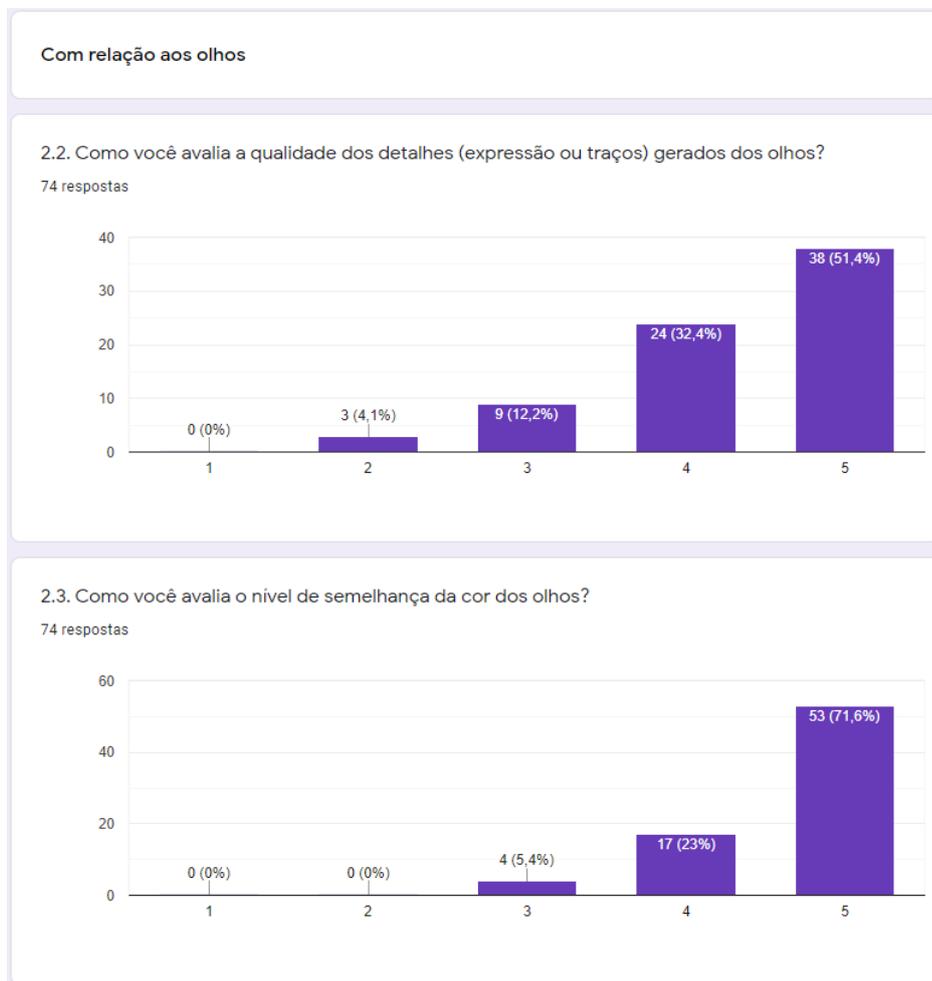
Fonte: Própria do autor.

Figura 83 – Resultados - Características Faciais - Will Smith - parte 1



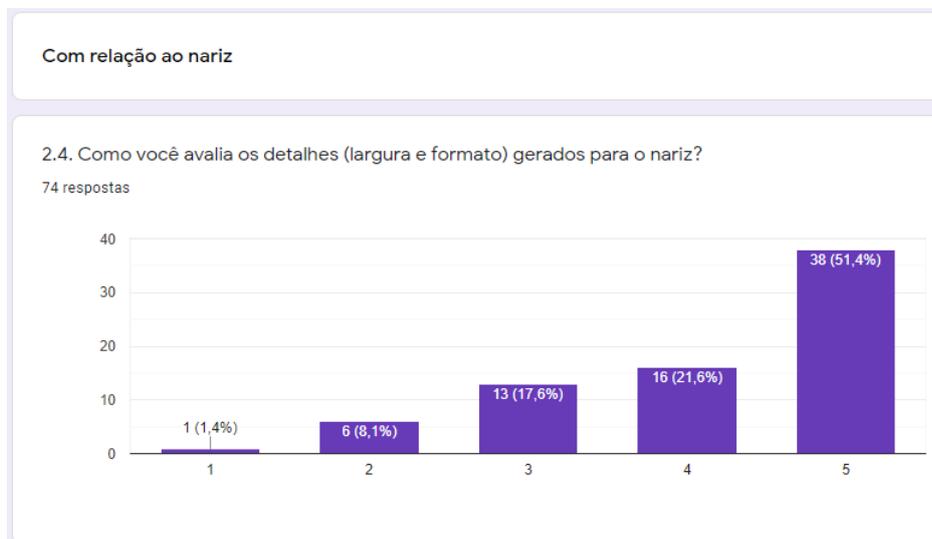
Fonte: Própria do autor.

Figura 84 – Resultados - Características Faciais - Will Smith - parte 2



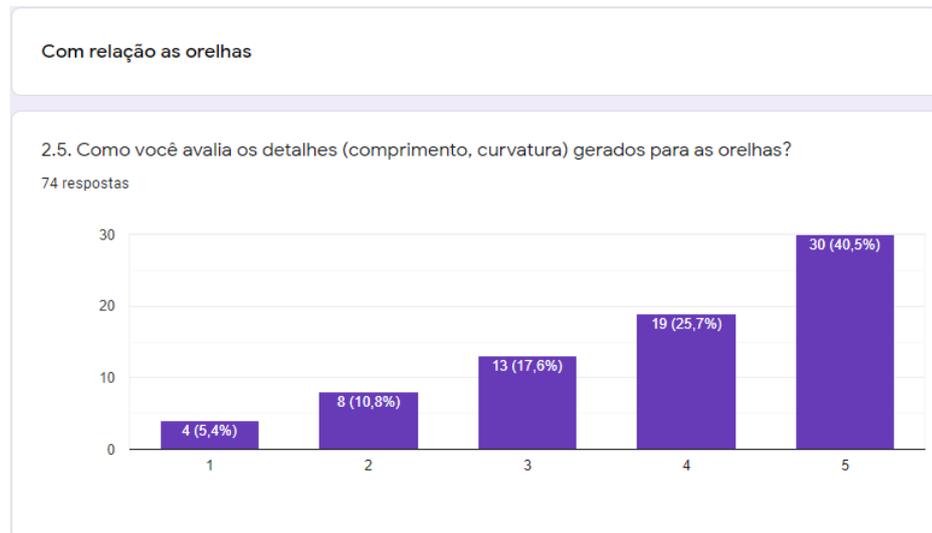
Fonte: Própria do autor.

Figura 85 – Resultados - Características Faciais - Will Smith - parte 3



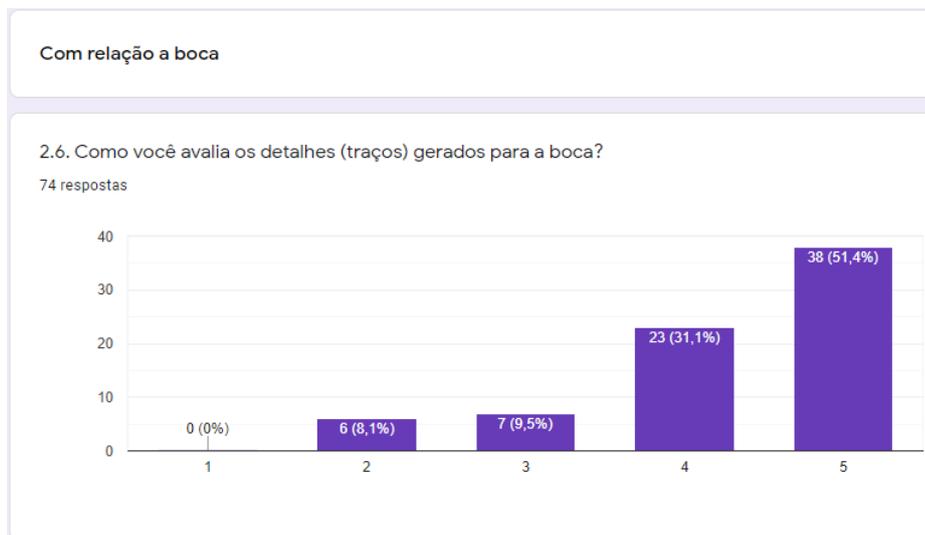
Fonte: Própria do autor.

Figura 86 – Resultados - Características Faciais - Will Smith - parte 4



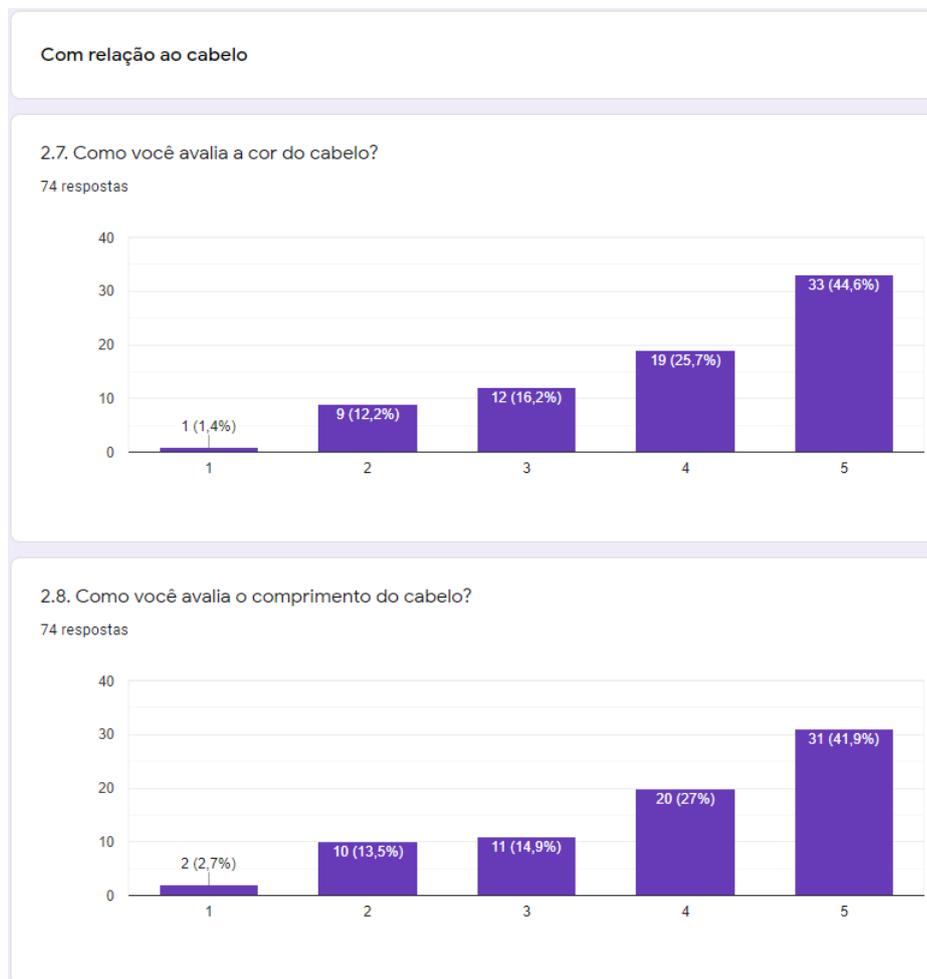
Fonte: Própria do autor.

Figura 87 – Resultados - Características Faciais - Will Smith - parte 5



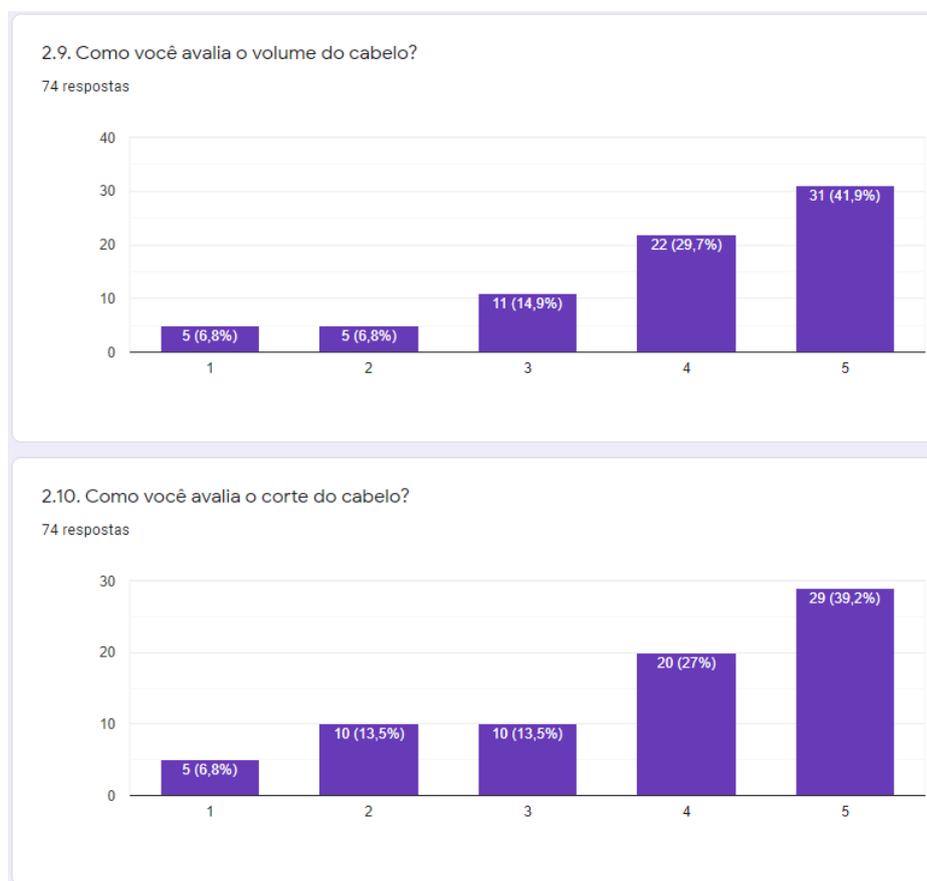
Fonte: Própria do autor.

Figura 88 – Resultados - Características Faciais - Will Smith - parte 6



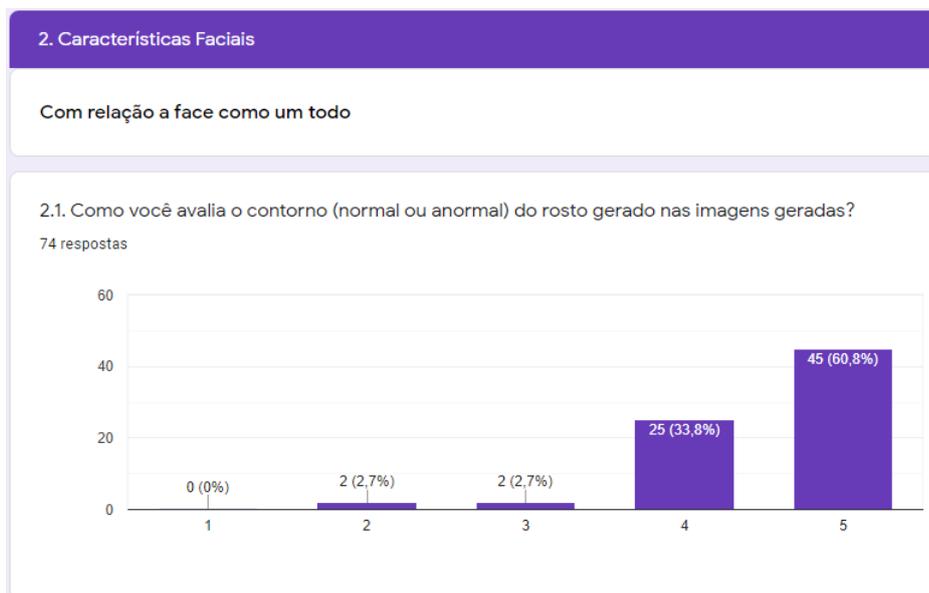
Fonte: Própria do autor.

Figura 89 – Resultados - Características Faciais - Will Smith - parte 7



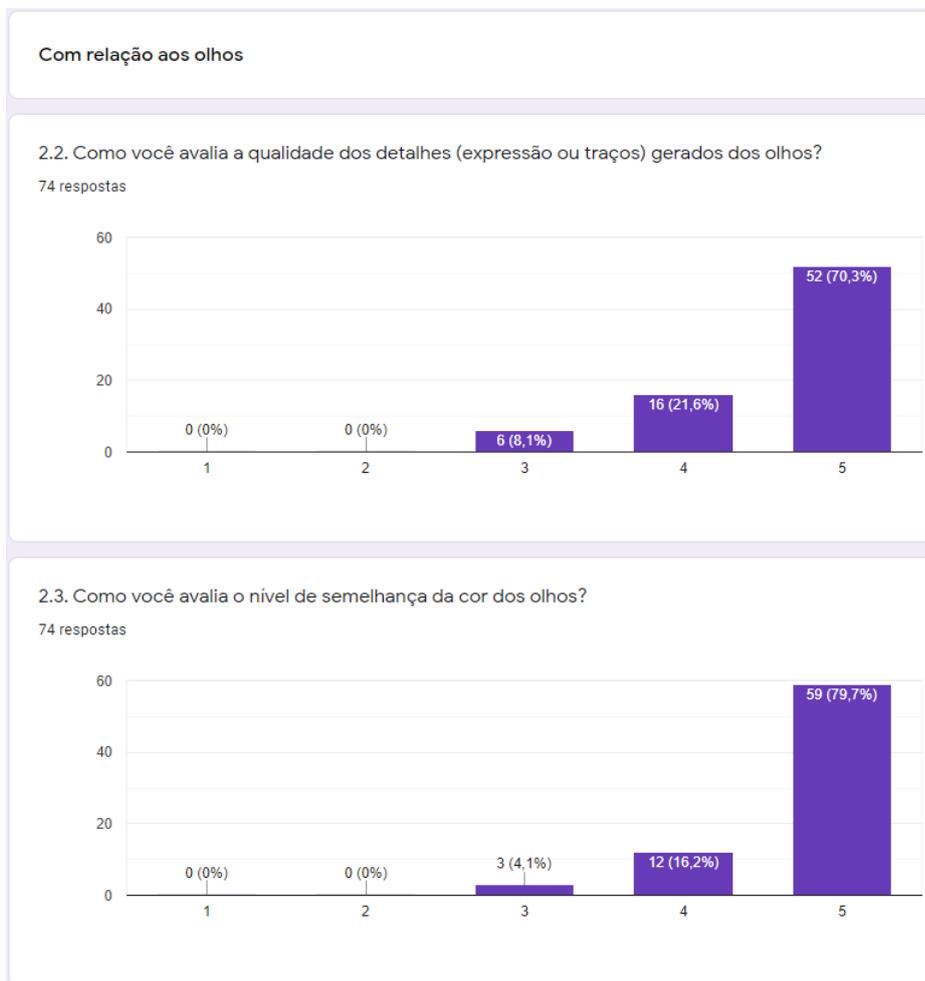
Fonte: Própria do autor.

Figura 90 – Resultados - Características Faciais - Eunice Baía - parte 1



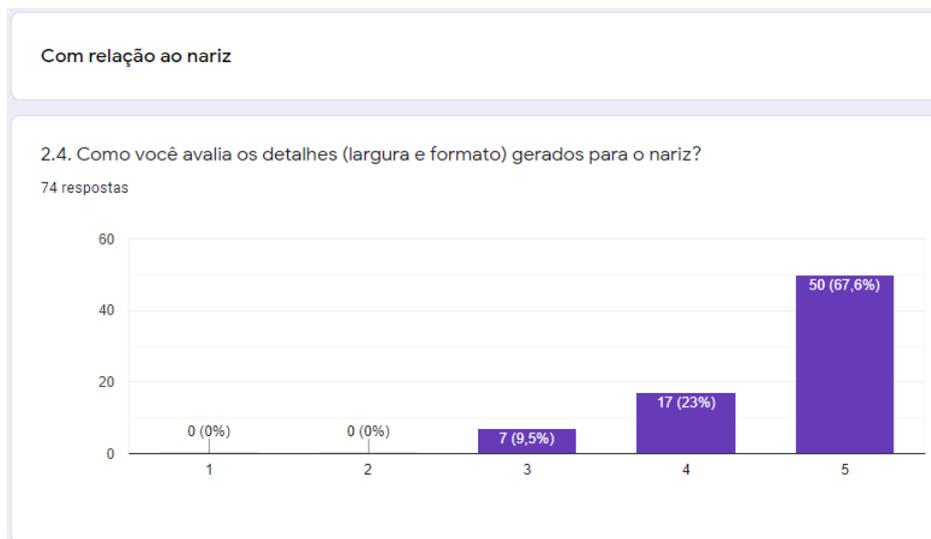
Fonte: Própria do autor.

Figura 91 – Resultados - Características Faciais - Eunice Baía - parte 2



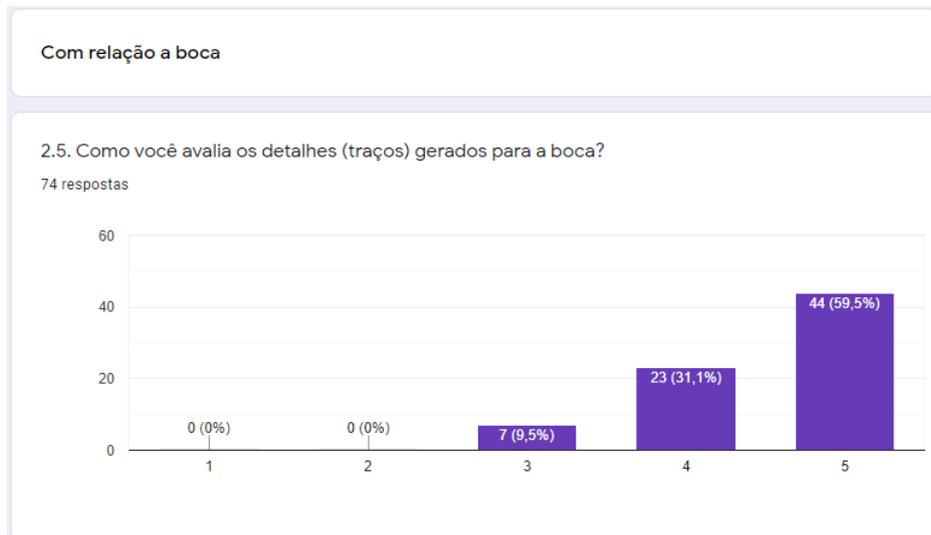
Fonte: Própria do autor.

Figura 92 – Resultados - Características Faciais - Eunice Baía - parte 3



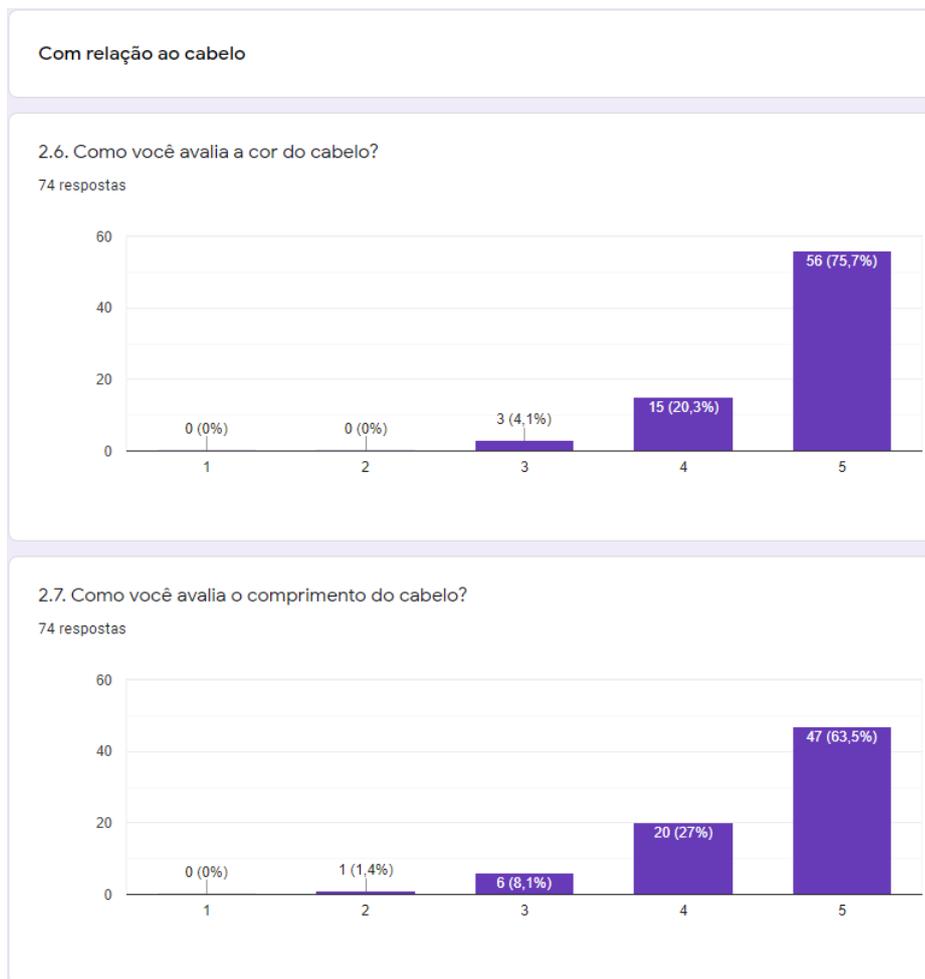
Fonte: Própria do autor.

Figura 93 – Resultados - Características Faciais - Eunice Baía - parte 4



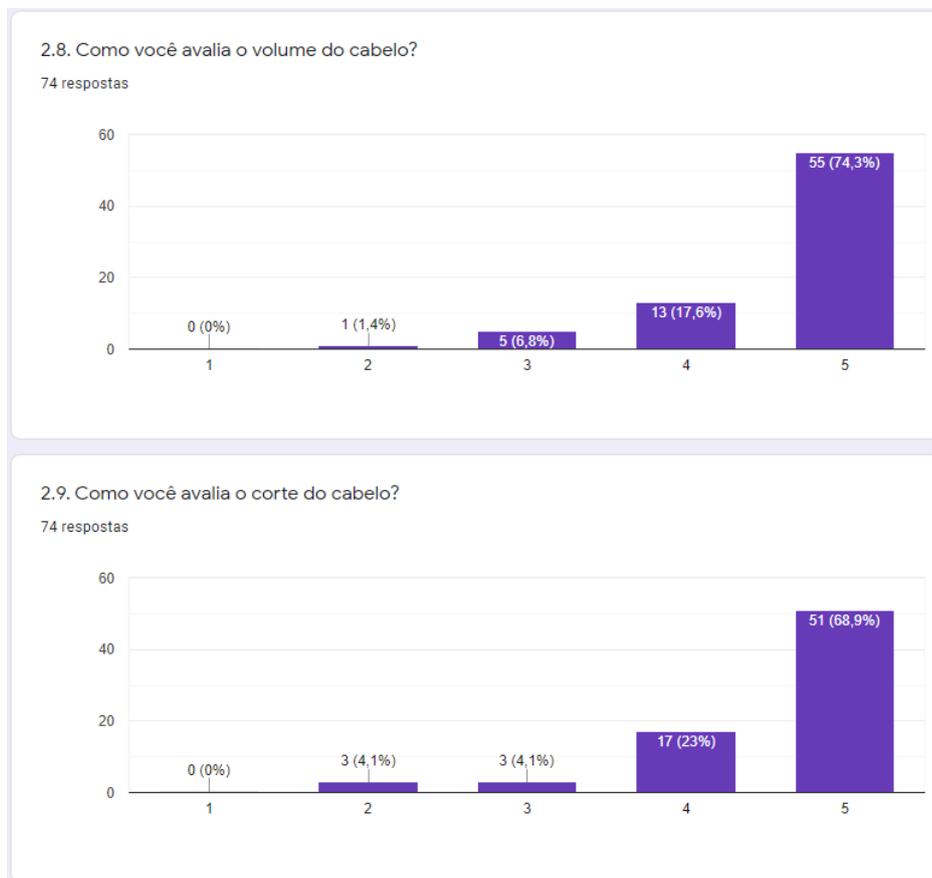
Fonte: Própria do autor.

Figura 94 – Resultados - Características Faciais - Eunice Baía - parte 5



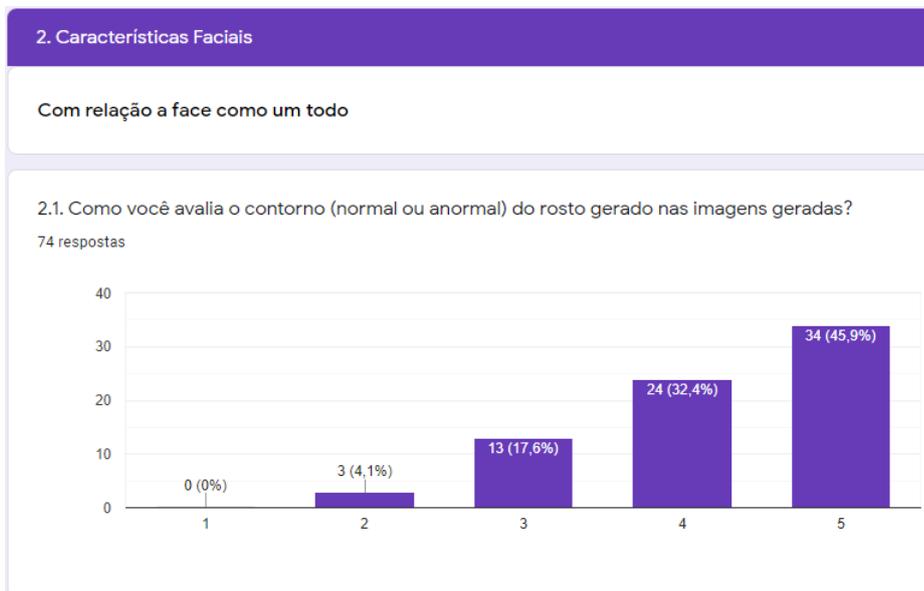
Fonte: Própria do autor.

Figura 95 – Resultados - Características Faciais - Eunice Baía - parte 6



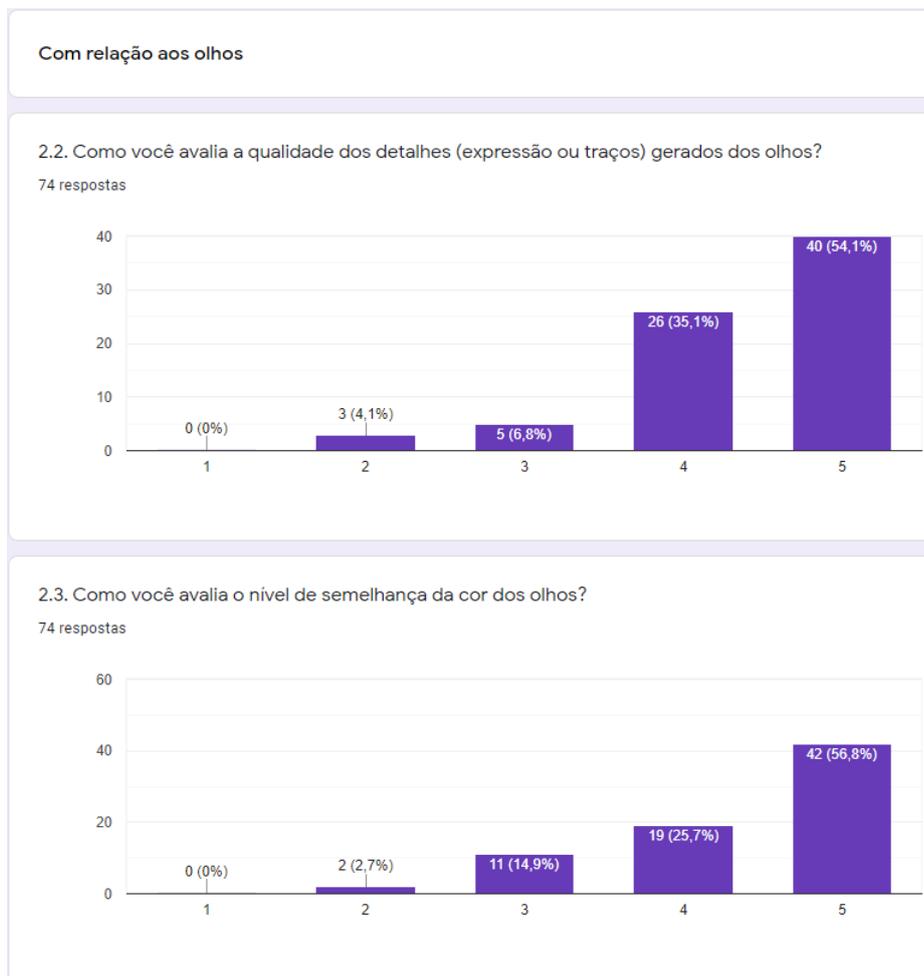
Fonte: Própria do autor.

Figura 96 – Resultados - Características Faciais - Daniele Suzuki - parte 1



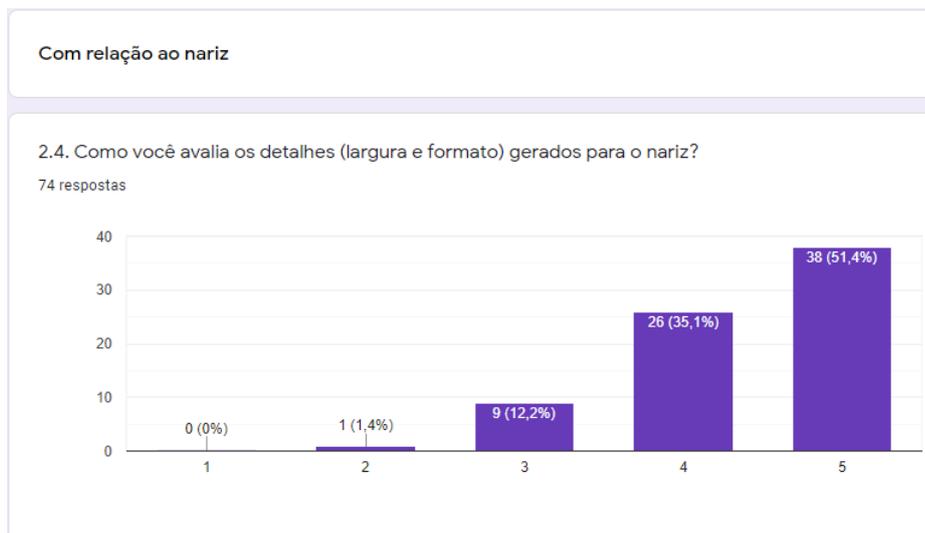
Fonte: Própria do autor.

Figura 97 – Resultados - Características Faciais - Daniele Suzuki - parte 2



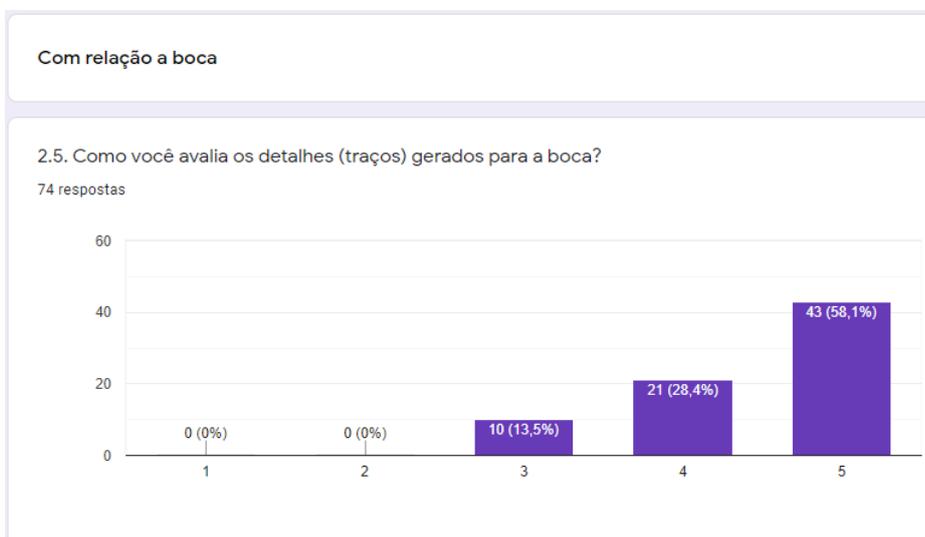
Fonte: Própria do autor.

Figura 98 – Resultados - Características Faciais - Daniele Suzuki - parte 3



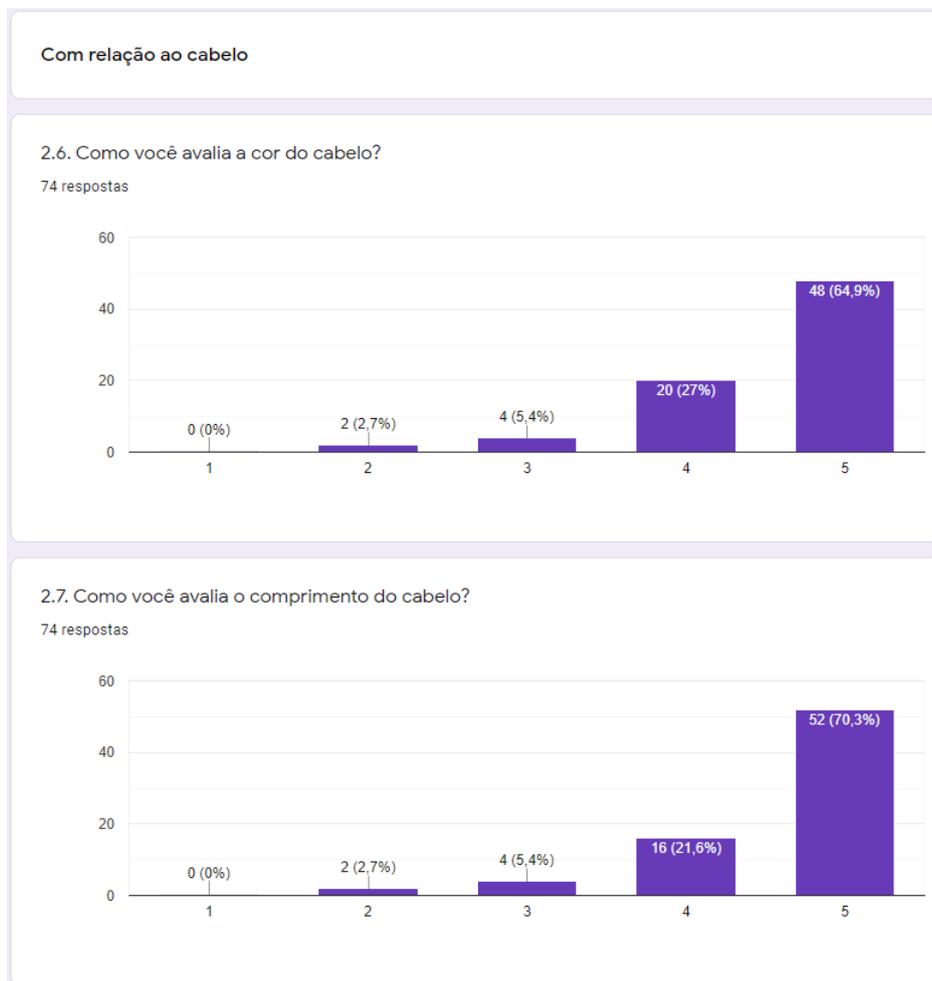
Fonte: Própria do autor.

Figura 99 – Resultados - Características Faciais - Daniele Suzuki - parte 4



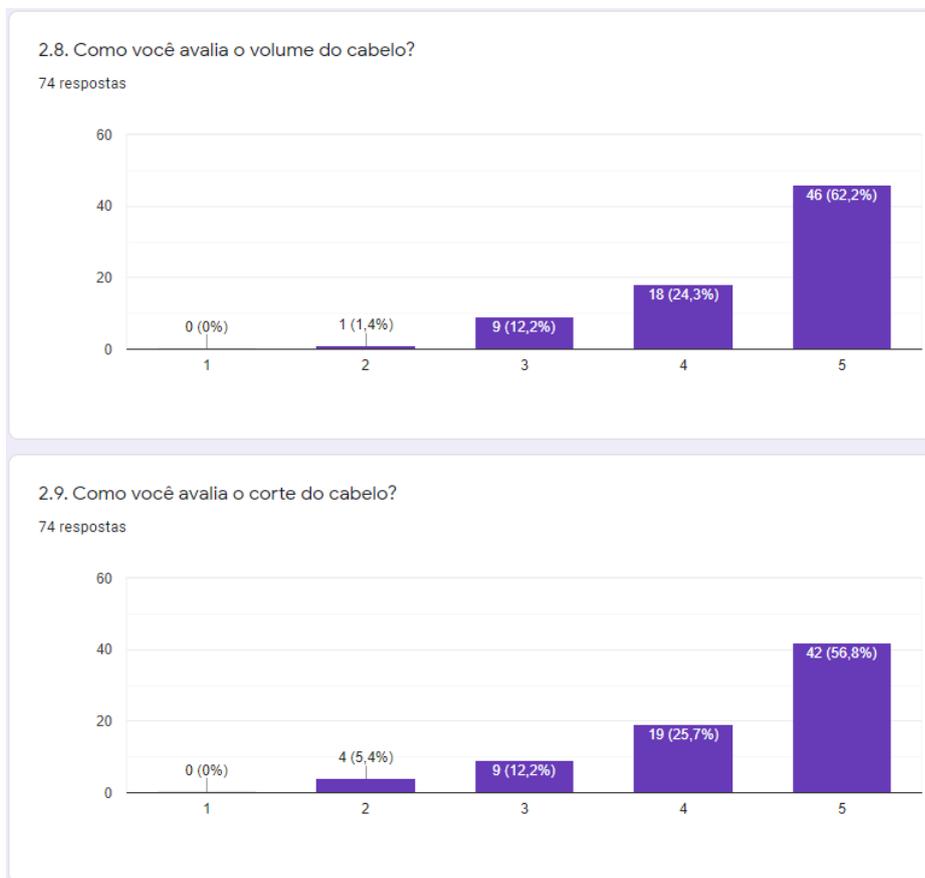
Fonte: Própria do autor.

Figura 100 – Resultados - Características Faciais - Daniele Suzuki - parte 5



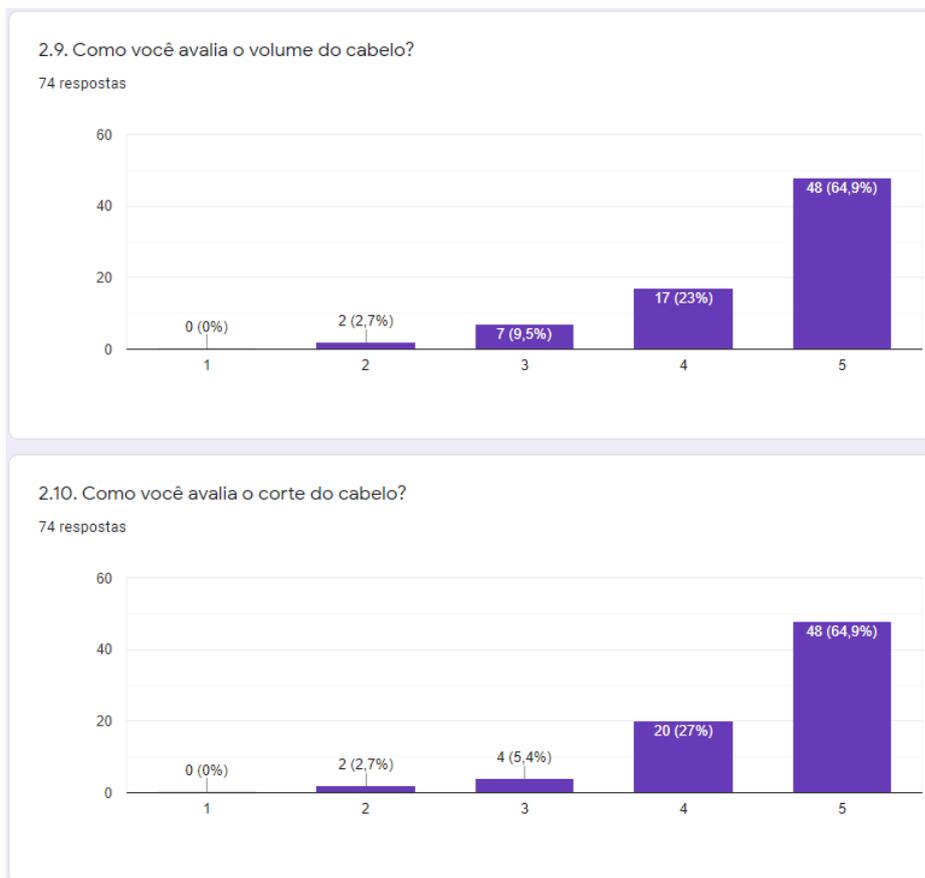
Fonte: Própria do autor.

Figura 101 – Resultados - Características Faciais - Daniele Suzuki - parte 6



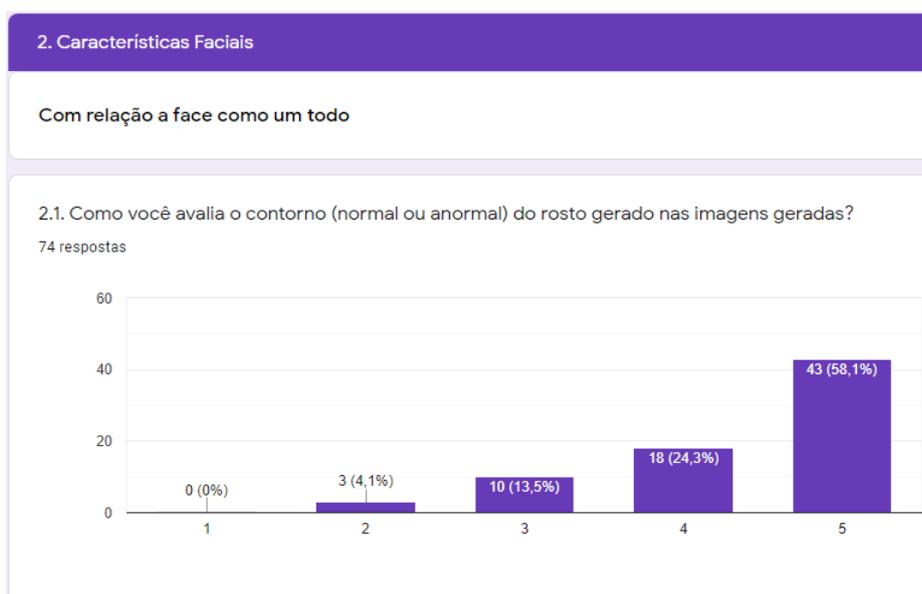
Fonte: Própria do autor.

Figura 102 – Resultados - Características Faciais - Daniele Suzuki - parte 7



Fonte: Própria do autor.

Figura 103 – Resultados - Características Faciais - Leonardo Dicaprio - parte 1



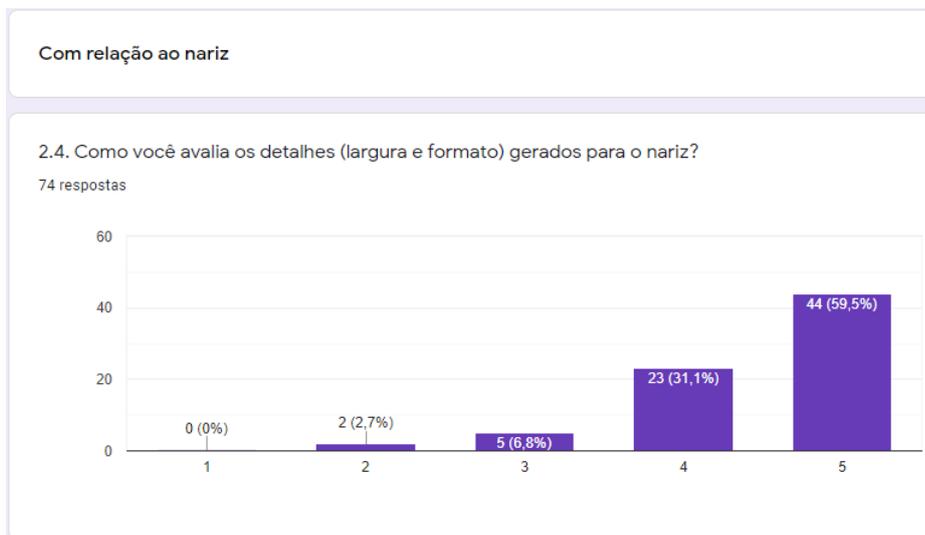
Fonte: Própria do autor.

Figura 104 – Resultados - Características Faciais - Leonardo Dicaprio - parte 2



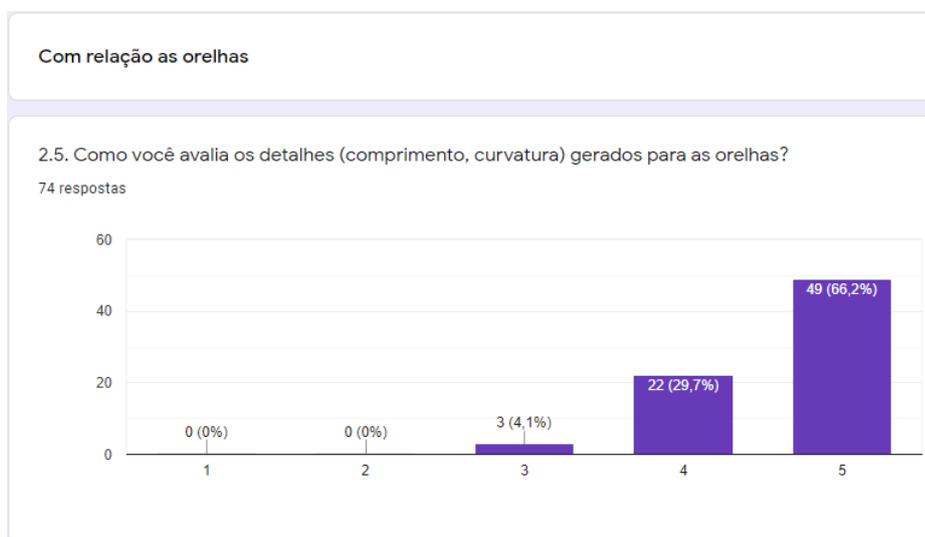
Fonte: Própria do autor.

Figura 105 – Resultados - Características Faciais - Leonardo Dicaprio - parte 3



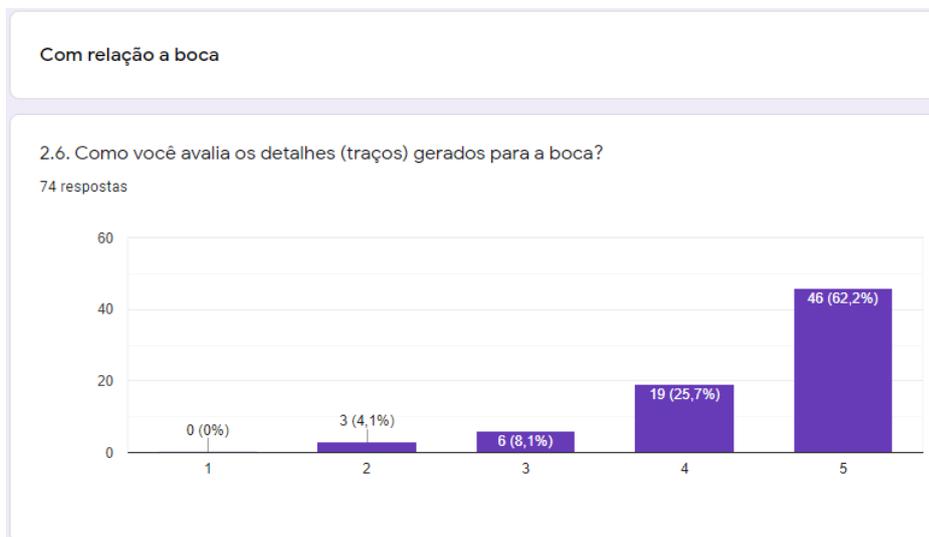
Fonte: Própria do autor.

Figura 106 – Resultados - Características Faciais - Leonardo Dicaprio - parte 4



Fonte: Própria do autor.

Figura 107 – Resultados - Características Faciais - Leonardo Dicaprio - parte 5



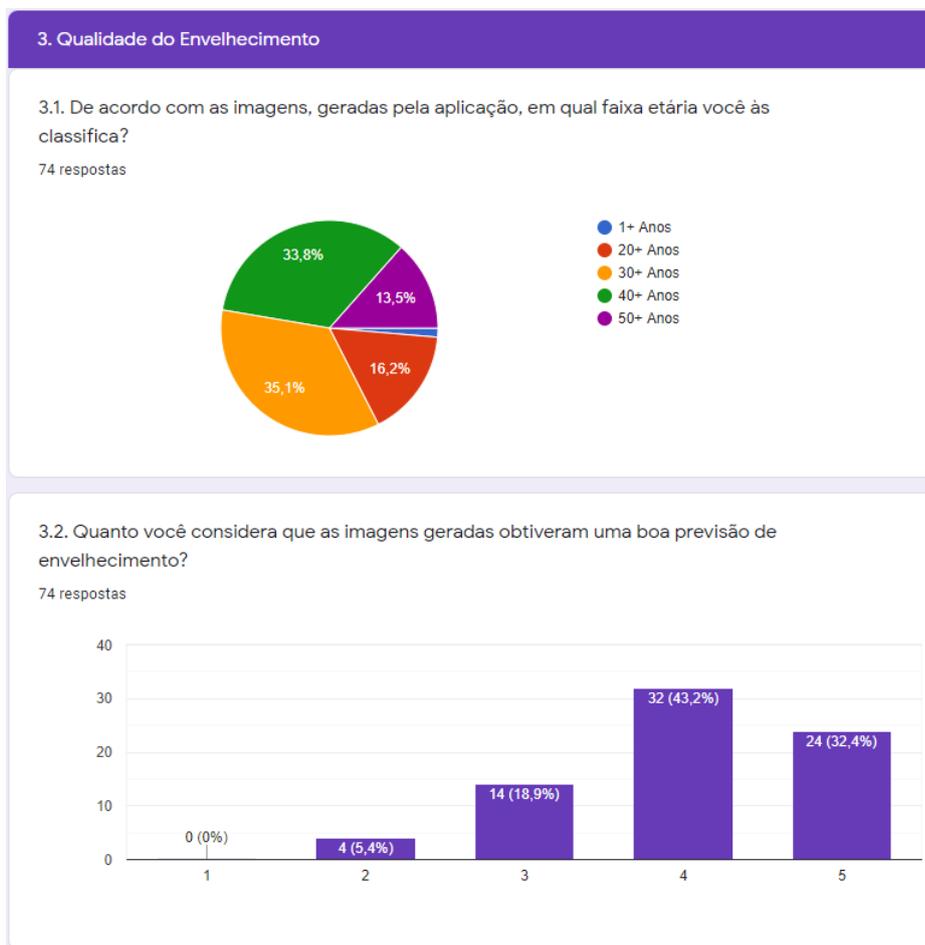
Fonte: Própria do autor.

Figura 108 – Resultados - Características Faciais - Leonardo Dicaprio - parte 6



Fonte: Própria do autor.

Figura 109 – Resultados - Qualidade do envelhecimento - parte 1



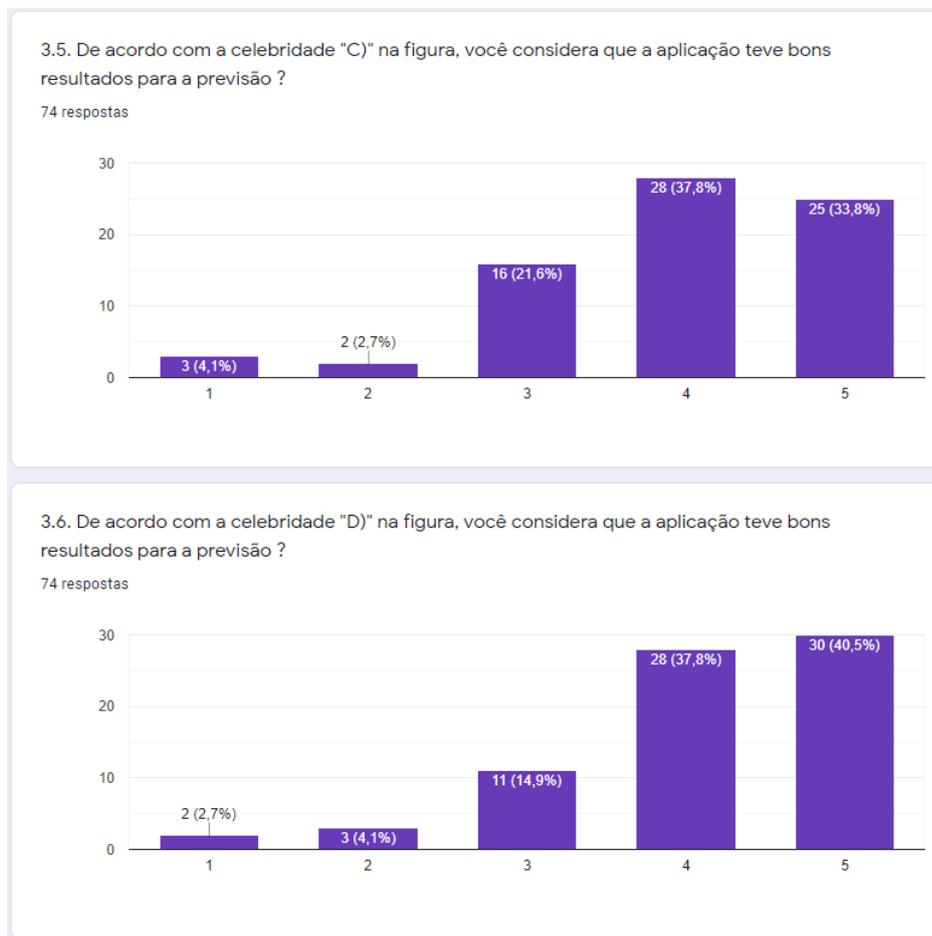
Fonte: Própria do autor.

Figura 110 – Resultados - Qualidade do envelhecimento - parte 2



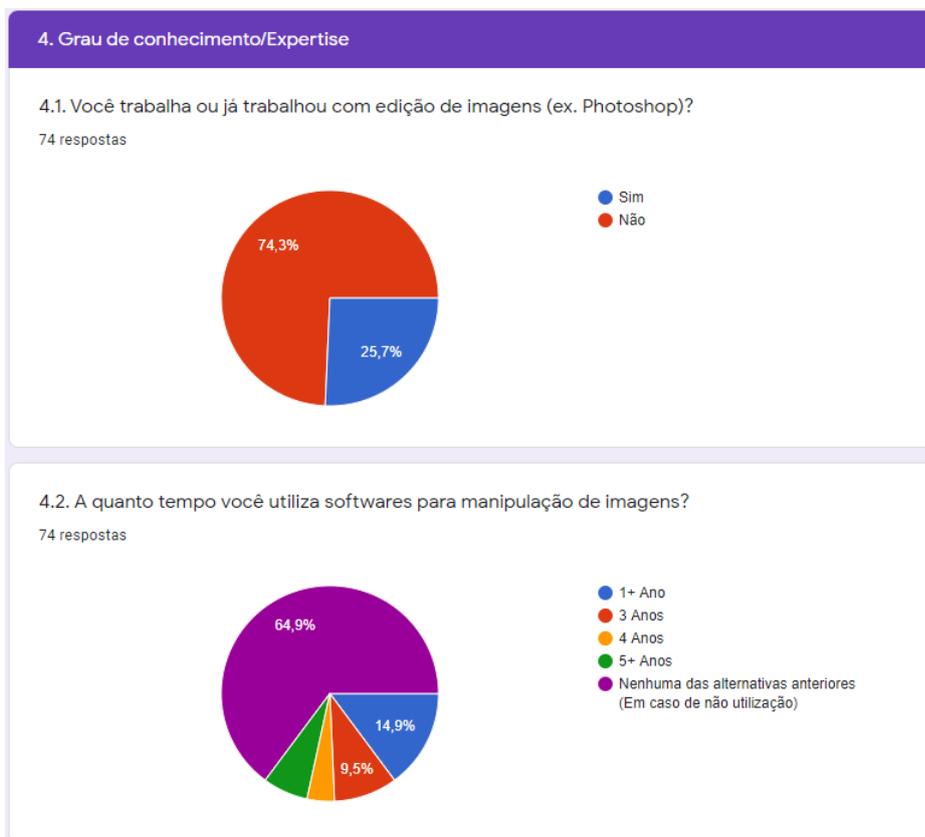
Fonte: Própria do autor.

Figura 111 – Resultados - Qualidade do envelhecimento - parte 3



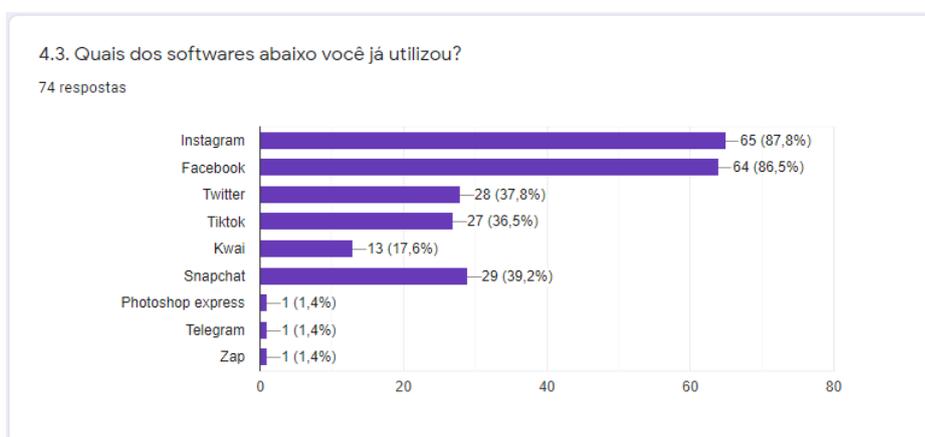
Fonte: Própria do autor.

Figura 112 – Resultados - Grau de conhecimento/Expertise - parte 1



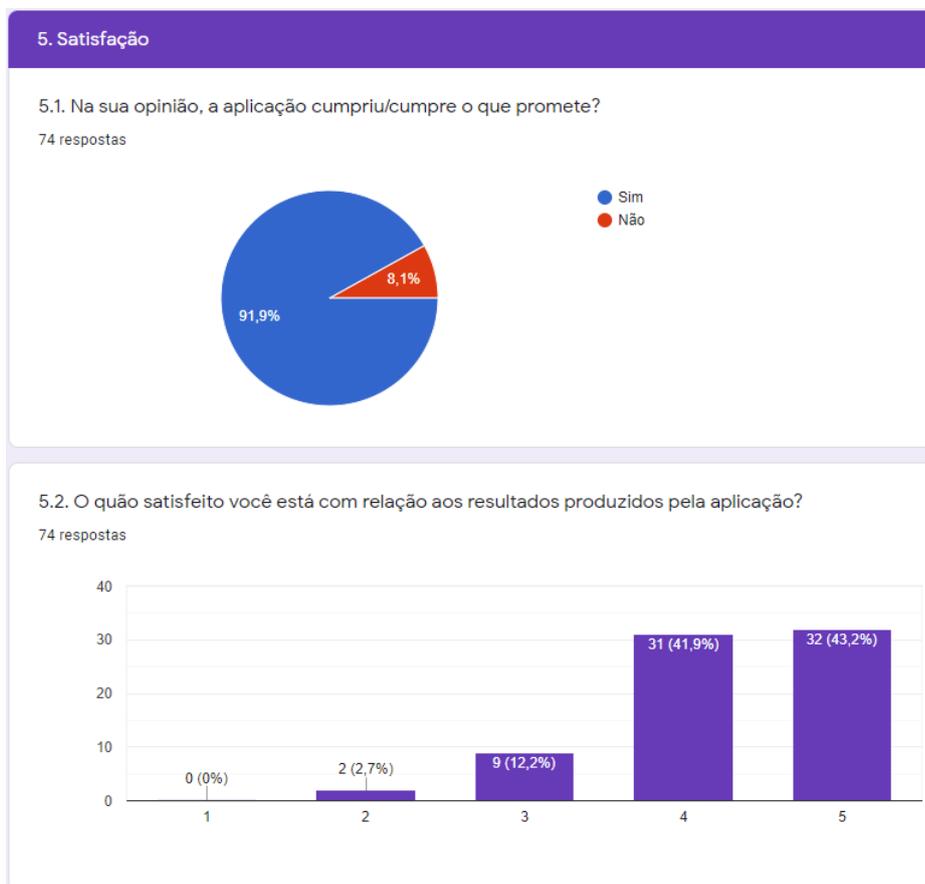
Fonte: Própria do autor.

Figura 113 – Resultados - Grau de conhecimento/Expertise - parte 2



Fonte: Própria do autor.

Figura 114 – Resultados - Satisfação



Fonte: Própria do autor.