

SmileDataAI: segmentação semântica de dentes em imagens de raio-x por meio de aprendizado profundo.

Vinícius Massaru Oliveira Tanaka
Universidade Federal de São Paulo - UNIFESP
São José dos Campos - São Paulo, Brasil
massaru.tanaka@unifesp.br

Resumo—A segmentação semântica por meio de aprendizado profundo é uma área em crescimento no processamento de imagens. Neste contexto, a odontologia se beneficia significativamente ao explorar o poder do processamento de imagens aliado ao uso de técnicas de aprendizado profundo. Este artigo apresenta uma abordagem de segmentação semântica baseada no modelo HRNet, aplicada a imagens radiográficas odontológicas. O objetivo principal é desenvolver um modelo de rede neural capaz de gerar máscaras nas imagens de raio-x, visando a classificação precisa de cada dente. Ao combinar as capacidades de segmentação semântica e aprendizado profundo, espera-se melhorar o diagnóstico odontológico, fornecendo informações valiosas para a tomada de decisões clínicas e personalizando o tratamento odontológico.

Index Terms—machine learning, segmentação semântica, odontologia, inteligência artificial, redes neurais artificiais, HRNet, deep learning, redes neurais convolucionais

I. INTRODUÇÃO E MOTIVAÇÃO

A área de processamento de imagens tem se desenvolvido muito com auxílio de Inteligência Artificial (IA), particularmente em conjunto com ferramentas e modelos de Machine Learning (ML). Essa combinação de tecnologias tem mostrado um potencial significativo na identificação de objetos e classificação de elementos orofaciais [2] [6].

Para a odontologia, essa sinergia é uma ferramenta poderosa, pois possibilita uma análise mais precisa e eficiente das imagens radiográficas. Ao explorar o poder do processamento de imagens aliado com o uso de ML, torna-se viável aprimorar o diagnóstico odontológico, proporcionando uma base sólida para a tomada de decisões clínicas. Neste contexto, este artigo explora as possibilidades e os benefícios dessa combinação de tecnologias na odontologia, destacando como ela pode melhorar a identificação e classificação de dentes em imagens radiográficas, contribuindo para um tratamento odontológico mais preciso e personalizado [3].

II. CONCEITOS FUNDAMENTAIS E TRABALHOS RELACIONADOS

A. Inteligência Artificial e Machine Learning

Podemos dizer que a inteligência artificial (IA) é um campo da ciência da computação que se concentra no desenvolvimento de sistemas capazes de realizar tarefas que normalmente requerem inteligência humana (Russell e Norvig,

1995, Capítulo 1). Esses sistemas são projetados para imitar, simular ou reproduzir características da inteligência humana. Para este artigo é necessário entender os conceitos de machine learning como Aprendizado Supervisionado, Aprendizado Não Supervisionado que são abordados por Russel e Norvig em seu livro “Inteligência Artificial” no capítulo 18 [1].

Afinal, o que é *Machine Learning*? O aprendizado de máquina (traduzido para o português) é um ramo da Inteligência Artificial que se concentra no desenvolvimento de algoritmos e modelos capazes de permitir que os sistemas “aprendam” a partir de dados e experiências passadas. Esses sistemas são projetados para identificar padrões, fazer previsões ou tomar decisões com base nos exemplos ou informações fornecidas. O aprendizado de máquina é alimentado por um conjunto diversificado de algoritmos, que são treinados com dados relevantes para melhorar seu desempenho ao longo do tempo [1].

Na literatura, o aprendizado supervisionado é uma abordagem no aprendizado de máquina em que um conjunto de dados de treinamento é fornecido, onde cada exemplo é acompanhado por uma etiqueta ou rótulo conhecido. O objetivo do aprendizado supervisionado é ensinar ao modelo a relação entre os recursos ou características dos dados de entrada e suas respectivas saídas esperadas. O modelo é treinado para generalizar essas relações e ser capaz de fazer previsões precisas em novos dados, onde as saídas esperadas são desconhecidas. Exemplos de algoritmos de aprendizado supervisionado incluem regressão linear, árvores de decisão e máquinas de vetores de suporte [1].

Por outro lado, o aprendizado não supervisionado é outra abordagem no aprendizado de máquina, onde não há rótulos ou saídas esperadas fornecidas nos dados de treinamento. Em vez disso, o objetivo é encontrar estruturas ocultas, padrões ou agrupamentos nos dados sem nenhuma informação prévia. Os algoritmos de aprendizado não supervisionado exploram a estrutura intrínseca dos dados para organizar, agrupar ou reduzir a dimensionalidade dos mesmos. Isso pode ajudar na descoberta de informações úteis, segmentação de dados ou identificação de anomalias. Exemplos de algoritmos de aprendizado não supervisionado incluem clusterização, análise de componentes principais e regras de associação [1].

Ao explorar o aprendizado de máquina, tanto supervisio-

nado quanto não supervisionado, pode-se alcançar avanços significativos em várias aplicações. No entanto, um dos enfoques mais promissores para o artigo é o uso de Redes Neurais Artificiais. As redes neurais são modelos inspirados no funcionamento do cérebro humano, compostos por camadas de neurônios interconectados. Essas redes têm a capacidade de aprender padrões complexos, extraír características relevantes dos dados e realizar, por exemplo, o processamento de imagens que é um dos objetivos do artigo. Além disso, as redes neurais possuem a capacidade de realizar as tarefas de classificação, regressão, processamento de linguagem natural, e muito mais. O próximo tópico deste artigo se aprofundará nas Redes Neurais Artificiais e suas aplicações no campo da Inteligência Artificial [3].

B. Redes Neurais Artificiais

O que podemos chamar de rede neural na computação?

Uma rede neural é um modelo computacional inspirado no funcionamento do cérebro humano, composto por um conjunto de unidades interconectadas, chamadas neurônios artificiais ou unidades de processamento, que trabalham em conjunto para realizar tarefas de aprendizado e processamento de informações. Cada neurônio artificial recebe entradas ponderadas, realiza uma soma ponderada dessas entradas e aplica uma função de ativação para determinar sua saída. As redes neurais podem ser organizadas em várias camadas, como a camada de entrada, camadas ocultas e a camada de saída, permitindo a modelagem de relações complexas entre os dados [1] [4] [5].

O primeiro neurônio de uma rede neural artificial criado por Rosenblatt em 1958 e foi chamada de *perceptron*. A perceptron foi desenvolvida com base em um modelo matemático desenvolvido por McCulloch e Pitts (1943), onde as entradas são ponderadas e somadas, e o resultado é passado por uma função de ativação conforme mostrado na “Fig. 1” que nos retorna uma saída que pode ser um resultado esperado ou outras conexões entre perceptrons. A função de ativação da perceptron geralmente é uma função degrau, que retorna um valor binário com base em um limiar predefinido [1] [4] [5].

E o que são as funções de ativação? As funções de ativação são cruciais para o funcionamento dos neurônios em redes neurais e são frequentemente definidas como:

$$a = g(z) \quad (1)$$

Em “(1)” a letra a é o cálculo da ativação, e a função $g(z)$ é a função de ativação definida pelo desenvolvedor. Elas determinam a saída de um neurônio com base na soma ponderada das entradas. As funções de ativação podem ser lineares ou não lineares [1] [4].

Segue abaixo algumas funções de ativação comumente utilizadas e seus respectivos gráficos:

- Função Sigmoid, que mapeia os valores de entrada para um intervalo entre 0 e 1 conforme “Fig. 2”. Uma função comumente utilizada para tarefas de classificação binária.
- Função Tangente Hiperbólica, que mapeia os valores para um intervalo entre -1 e 1 conforme “Fig. 3”. Esta

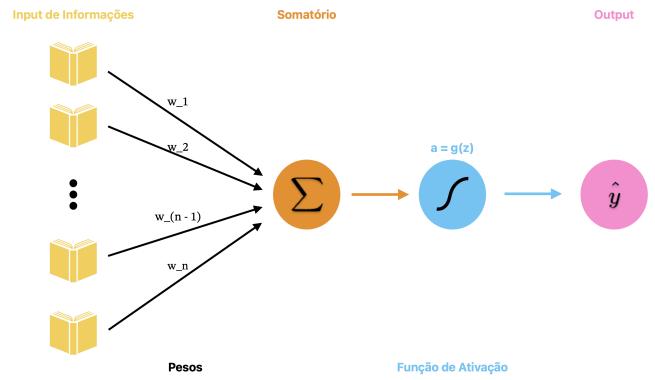


Figura 1. Comportamento de uma perceptron, no qual o neurônio artificial recebe informações que possuem pesos, realiza o somatório destes inputs, propaga as informações para a função de ativação que retorna uma saída de acordo com esta função.

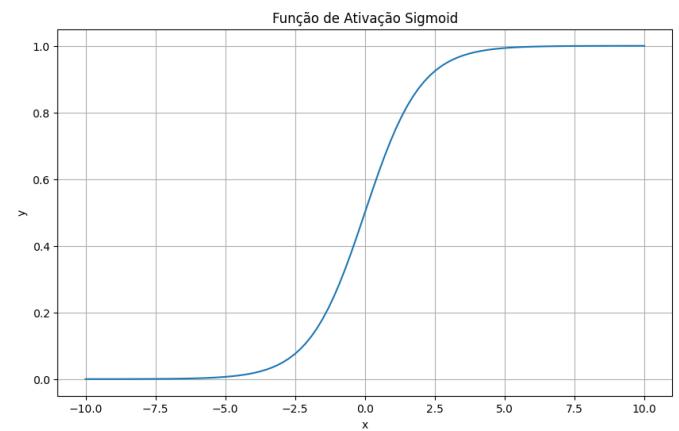


Figura 2. Gráfico representando o comportamento da função sigmoid que mapeia valores de -10 até 10 no eixo x para os valores entre 0 e 1 no eixo y

função pode ser utilizada para tarefas de classificação multiclasse.

- Função ReLU (Rectified Linear Unit), que retorna 0 para valores negativos e o próprio valor de entrada para valores positivos conforme “Fig. 4”.

As funções de ativação não lineares são importantes para permitir que as redes neurais modelem relações complexas e não lineares entre os dados de entrada (Capítulo 25 [4]) (Capítulo 18.7.1 [1]).

C. Processamento de Imagens e Segmentação Semântica

Das maravilhas do mundo macro como às paisagens naturais e as imagens do vasto universo que conhecemos até às minúcias do universo microscópico, o mundo visual é uma fonte infinita de informações. Como seres humanos, somos dotados de uma habilidade impressionante para interpretar e extraírem significado dessas imagens, mas e se pudéssemos ensinar as máquinas a fazer o mesmo? Com o avanço tecnológico podemos fazer isso com o uso de processamento de

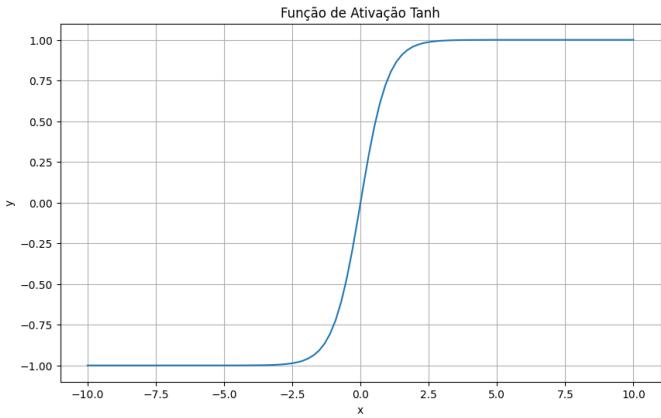


Figura 3. Gráfico representando o comportamento da função tangente hiperbólica, semelhante a função sigmoid possui um formato de "S", porém sua taxa de crescimento é maior.

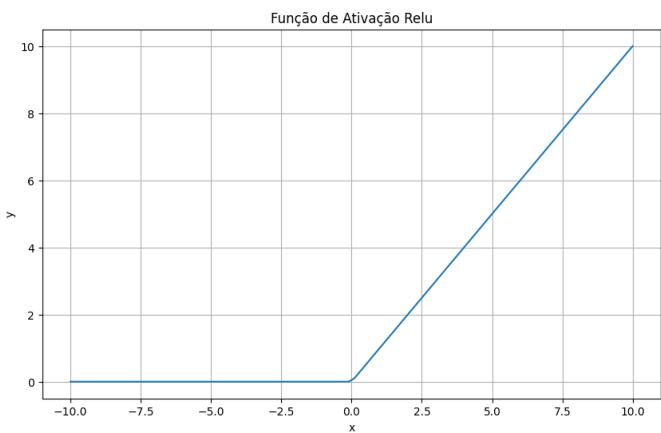


Figura 4. Gráfico representando o comportamento da função ReLU na qual mapeia valores de $x \leq 0$ como $y = 0$, e $x = y$ para valores em que $x > 0$.

imagens combinado ao aprendizado de máquina. *Afinal o que é o processamento de imagens?*

O processamento de imagens na inteligência artificial é um campo que visa capacitar as máquinas a compreender e interpretar o mundo visual ao nosso redor. É uma disciplina multidisciplinar que combina conceitos de visão computacional e aprendizado de máquina para extrair informações valiosas de imagens digitais. Ele desempenha um papel crucial em várias aplicações, desde reconhecimento facial, detecção de objetos, diagnósticos médicos e/ou odontológicos, até mesmo em desenvolvimento de veículos autônomos [7].

As seguintes etapas são utilizadas para elaborar o processamento de imagens:

- **Aquisição de Imagem:** aqui precisamos obter a imagem de forma digitalizada, que pode ser uma imagem de câmera fotográfica, escaneada e até mesmo de outras fontes. A qualidade da imagem adquirida desempenha um papel crítico na precisão e no desempenho do processamento subsequente. Neste artigo serão utilizadas imagens de raio-x panorâmica.

- **Pré-processamento:** antes de aplicar algum algoritmo específico é necessário realizar um pré-processamento das imagens para torná-las mais adequadas para o modelo de máquina. Isso pode incluir etapas como redimensionamento, normalização, correção de cores, remoção de ruídos e realce de características relevantes.

- **Extração de Características:** nesta etapa, são identificadas características relevantes na imagem que podem ser usadas para descrever e representar seu conteúdo. Isso pode incluir a detecção de bordas, extração de texturas, identificação de formas e padrões, entre outras técnicas. A extração de recursos é fundamental para fornecer informações discriminativas e significativas sobre a imagem.

- **Representação dos Dados:** após a extração de recursos, os dados da imagem são representados de uma forma que os algoritmos de aprendizado de máquina possam processar. Geralmente, isso envolve a transformação das características extraídas em uma forma numérica, como vetores ou tensores, que podem ser alimentados em algoritmos de treinamento e inferência.

- **Escolha de um Modelo de Aprendizado:** os algoritmos de aprendizado de máquina são aplicados aos dados da imagem para aprender padrões e relacionamentos entre os recursos extraídos. As redes neurais convolucionais (CNNs) são uma das arquiteturas de aprendizado profundo mais populares para processamento de imagens, permitindo que as máquinas aprendam automaticamente a partir de grandes conjuntos de dados rotulados.

- **Análise e interpretação de dados:** Com o modelo treinado, a máquina pode interpretar e analisar novas imagens. Isso pode envolver tarefas como reconhecimento de objetos, classificação de imagens, detecção de características específicas e muito mais. A máquina pode tomar decisões ou fornecer insights com base nas informações extraídas da imagem.

Para prosseguir com o artigo precisamos entender o que seria *segmentação semântica*. Ela é uma técnica de visão computacional que permite identificar e segmentar diferentes objetos e regiões em uma imagem com base no seu significado e contexto. Em vez de apenas detectar objetos, a segmentação semântica vai além, atribuindo rótulos a cada pixel na imagem, criando assim uma máscara que separa visualmente cada objeto ou região. Essa abordagem fornece uma compreensão mais detalhada da cena, permitindo que as máquinas entendam a estrutura e a semântica do ambiente visual. A segmentação semântica tem uma ampla gama de aplicações, incluindo aperfeiçoamento de sistemas de condução autônoma, análise de imagens médicas e realidade aumentada, proporcionando um avanço significativo na capacidade de interpretar e interagir com o mundo visualmente.

D. Métricas Para Avaliação de Aprendizado em Segmentação Semântica

Pixel Accuracy (Precisão de Pixel) é uma métrica simples e direta que mede a proporção de pixels corretamente classifi-

cados em relação ao total de pixels na imagem. Essa métrica fornece uma medida de quão precisamente o modelo de segmentação semântica consegue atribuir os rótulos corretos para cada pixel da imagem. No entanto, a Pixel Accuracy pode ser enganosa se houver classes desbalanceadas, pois um modelo pode alcançar uma alta precisão simplesmente classificando a maioria dos pixels como a classe dominante, ignorando outras classes menos frequentes [9].

IoU (Intersection over Union ou Jaccard Index) é uma métrica que calcula a sobreposição entre a máscara prevista pelo modelo e a máscara de referência (rótulo verdadeiro). O IoU é calculado dividindo a área de sobreposição entre as duas máscaras pela área da sua união conforme a equação “(2)”:

$$\text{IoU} = \frac{\text{Área de Sobreposição}}{\text{Área de União}} \quad (2)$$

O IoU varia entre 0 e 1, sendo que 1 indica uma segmentação perfeita e 0 indica uma segmentação completamente incorreta. O IoU é uma métrica amplamente usada para avaliar a qualidade da segmentação semântica, especialmente quando as classes estão desbalanceadas [9] [10].

O Dice Coefficient, também conhecido como F1 Score, é outra métrica comumente usada para avaliar a qualidade da segmentação semântica. É calculado como o dobro da interseção entre a máscara prevista e a máscara de referência, dividido pela soma das áreas das duas máscaras conforme a equação “(3)”:

$$\text{F1 Score} = \frac{2 \times \text{Área de Sobreposição}}{\text{Total de pixels das imagens}} \quad (3)$$

O Dice Coefficient varia entre 0 e 1, onde 1 representa uma segmentação perfeita e 0 representa uma segmentação totalmente incorreta. Assim como o IoU, o Dice Coefficient é útil para lidar com classes desbalanceadas, pois leva em consideração a sobreposição entre as máscaras [9].

E. Anatomia e Notação Dentária

A anatomia dentária é o estudo da estrutura e características dos dentes humanos. Os dentes são estruturas duras e calcificadas que se encontram na boca e desempenham papéis essenciais na mastigação, na fala e na estética facial. Os dentes são classificados em diferentes tipos, cada um com uma forma e função específicas. Compreender a anatomia dentária é fundamental para a odontologia e para o cuidado adequado dos dentes e gengivas [8].

A notação dentária é um sistema de nomenclatura que atribui números e letras a cada dente da boca. Essa notação é utilizada para identificar e comunicar a localização específica de um dente dentro da cavidade oral. Existem duas notações dentárias amplamente utilizadas: a notação da Federação Dentária Internacional (FDI) e a notação de Palmer. Para este artigo usaremos a notação de Palmer [8].

A notação de Palmer utiliza letras maiúsculas para identificar os quadrantes da boca e números para identificar os dentes dentro de cada quadrante. A boca é dividida em quatro quadrantes: superior direito (Q1), superior esquerdo (Q2),

inferior esquerdo (Q3) e inferior direito (Q4). Os números são atribuídos de 1 a 8, começando do dente central incisivo até o terceiro molar. Por exemplo, o primeiro pré-molar superior direito seria representado como dente 4 na notação de Palmer (Q1-4) conforme as “Fig. 5” e “Fig. 6” [8].



Figura 5. Imagem retirada da página 15 [8]. Imagem bucal com visão frontal, mostrando os quadrantes de Q1 até Q4.

Na tabela I podemos identificar o quadrante do dente, sua posição, o número definido pela notação de Palmer e seu respectivo nome [8].

Tabela I
NOMENCLATURA DOS DENTES

Quadrante	Posição	Nºm.	Nome
Q1	1	11	Incisivo central superior direito
Q1	2	12	Incisivo lateral superior direito
Q1	3	13	Canino superior direito
Q1	4	14	Primeiro pré-molar superior direito
Q1	5	15	Segundo pré-molar superior direito
Q1	6	16	Primeiro molar superior direito
Q1	7	17	Segundo molar superior direito
Q1	8	18	Terceiro molar superior direito
Q2	1	21	Incisivo central superior esquerdo
Q2	2	22	Incisivo lateral superior esquerdo
...
Q3	1	31	Incisivo central inferior esquerdo
Q3	2	32	Incisivo lateral inferior esquerdo
...
Q4	1	41	Incisivo central inferior direito
Q4	2	42	Incisivo lateral inferior direito
...



Figura 6. Imagem retirada da página 17 [8]. Identificação dos elementos dentários de acordo com a numeração de 1 à 8 para cada quadrante Q1-4.

F. Trabalhos Relacionados

Nesta subseção, será apresentado uma visão geral dos trabalhos relacionados ao tema da segmentação semântica de imagens de raio-x. A literatura nessa área tem abordado diversos aspectos, desde técnicas tradicionais até abordagens baseadas em aprendizado profundo. A seguir, destacamos alguns trabalhos utilizados como referência para este artigo.

1) *Classification of Dental Radiographs Using Deep Learning*: Neste estudo, os autores avaliam retrospectivamente dados radiográficos e compararam o desempenho de três arquiteturas de aprendizado profundo para classificação. Os resultados obtidos mostram alta precisão em todas as arquiteturas, com destaque para a ResNet, que apresentou uma melhor performance, convergência rápida e menor variação. As características das imagens consideradas para classificação foram consistentes com o raciocínio de especialistas [11].

2) *Automatic segmentation of mandible in panoramic x-ray*: Resumidamente, neste estudo os autores buscam realizar a segmentação semântica de forma automática da mandíbula e da maxila de raio-x panorâmico. Foi a partir deste artigo que surgiu a ideia de realizar a segmentação semântica dos dentes e gerar um filtro que mostre todos os dentes de adultos contidos em um raio-x panorâmico. Além disso este trabalho foi o que forneceu a base de dados que será utilizada neste artigo. [6].

3) *Automated deep learning for classification of dental implant radiographs using a large multi-center dataset*: Este estudo busca realizar a classificação das marcas de parafusos de implante utilizados por meio de imagens de raio-x

panorâmica por meio de deep learning usando redes neurais convolucionais [12].

III. OBJETIVOS

Este artigo tem como objetivo principal criar um modelo de rede neural denominado *SmileDataAI* baseando-se no modelo HRNet que será abordado na seção *Metodologia Experimental* capaz de gerar máscaras nas imagens de raio-x no qual sejam classificados cada dente.

IV. METODOLOGIA EXPERIMENTAL

A. Tecnologias Utilizadas

1) *High-Resolution Network (HRNet)*: A HRNet, ou High-Resolution Network, é uma arquitetura de rede neural projetada para abordar o desafio de equilibrar a resolução espacial e a representação de recursos em tarefas de visão computacional conforme demonstrado sua arquitetura na “Fig. 7”. Ao contrário de outras arquiteturas convolucionais que agregam informações de várias resoluções em uma única via, a HRNet adota uma abordagem inovadora de manter representações de alta resolução em todas as etapas da rede. Isso é alcançado através da criação de várias vias paralelas, cada uma focada em uma determinada escala de resolução. Essas vias são então combinadas para formar uma representação final rica e detalhada. Essa abordagem preserva informações espaciais finas e permite que a HRNet capture características detalhadas em diferentes escalas, resultando em um desempenho aprimorado em tarefas como detecção de objetos, segmentação e pose estimation [13].

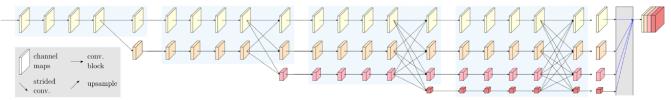


Figura 7. Representação da arquitetura HRNet. [15].

2) *Modelo de Rede Neural SegFormer*: A rede SegFormer é um modelo de rede neural projetado para resolver o problema de segmentação de imagens conforme a “Fig. 8”. A segmentação de imagens envolve dividir uma imagem em várias partes menores e rotulá-las de acordo com o objeto ou região que representam.

A arquitetura do SegFormer é baseada em uma rede neural chamada Transformer, que foi originalmente desenvolvida para tarefas de processamento de linguagem natural. No entanto, o SegFormer adaptou essa arquitetura para lidar com tarefas de segmentação de imagens.

O SegFormer começa recebendo uma imagem como entrada e a divide em pequenos pedaços chamados patches. Cada patch é então processado em uma etapa chamada codificação, onde as informações da imagem são extraídas e representadas em um formato mais compacto.

Em seguida, o SegFormer usa a arquitetura Transformer para analisar esses patches codificados e aprender padrões espaciais e contextuais na imagem. O Transformer usa mecanismos de atenção para capturar as relações entre os patches e entender como eles se relacionam uns com os outros.

Uma vez que o Transformer tenha analisado todos os patches, a etapa de decodificação ocorre. Nessa etapa, as informações codificadas são reconstruídas em uma representação espacial semelhante à imagem original. Essa reconstrução é feita por meio de um processo chamado up-sampling, que expande as representações de baixa resolução para correspondem à resolução original da imagem.

Finalmente, o SegFormer usa essas representações reconstruídas para rotular cada pixel da imagem com uma máscara de segmentação. Cada pixel é classificado de acordo com a classe de objeto ou região à qual pertence.

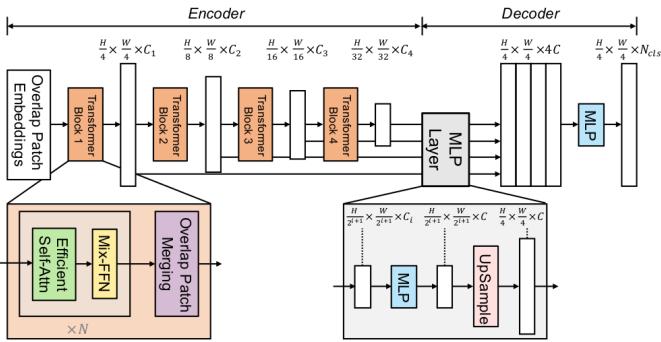


Figura 8. Representação da arquitetura SegFormer para tarefa de segmentação semântica [15].

3) **Arquitetura UPerNet:** A arquitetura utilizada no UPerNet (Unified Perceptual Parsing Network) consiste em uma combinação da Feature Pyramid Network (FPN) com o Pyramid Pooling Module (PPM) do PSPNet. A FPN é empregada como uma rede piramidal de características para extrair informações em diferentes escalas semânticas. O módulo PPM é adicionado à última camada da rede base antes de alimentar o ramo top-down do FPN. Além disso, várias cabeças (heads) são anexadas às diferentes camadas da rede para realizar tarefas específicas. Podemos visualizar sua arquitetura na “Fig. 9”

Durante o treinamento do UPerNet, cada iteração do processo de aprendizado utiliza uma mini-batch composta por imagens provenientes de várias fontes e tarefas diferentes. Para lidar com a diminuição efetiva do tamanho da mini-batch para cada tarefa, é aplicada uma estratégia de amostragem aleatória baseada na escala de cada fonte de dados em cada iteração. Apenas o caminho relacionado à fonte de dados selecionada é atualizado, focando apenas nas informações necessárias para inferir os conceitos relacionados a essa fonte específica. Além disso, é importante ressaltar que as perdas (losses) para segmentação de objeto e classificação de material não são calculadas em áreas não rotuladas (unlabeled).

Dessa forma, a arquitetura UPerNet busca aproveitar informações em diferentes escalas semânticas e realizar múltiplas tarefas simultaneamente, aprimorando a segmentação semântica em imagens [16].

4) **Ferramentas de Ciência de Dados:** Além da utilização do modelo HRNet, para a elaboração do projeto serão utilizados as seguintes ferramentas:

- **Google Colab:** Jupyter Notebook ambientado em nuvem.

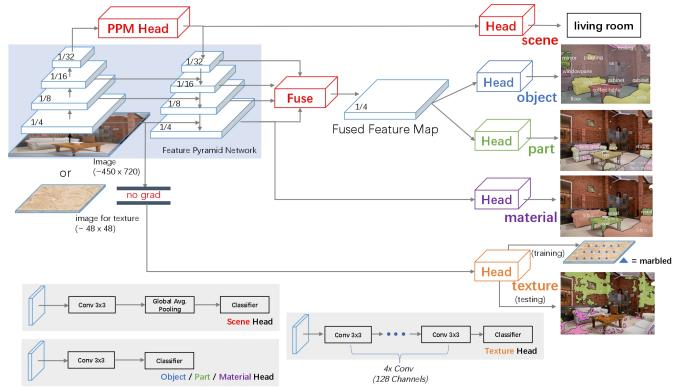


Figura 9. Representação da arquitetura UPerNet, para solução de tarefas multi-classe [16] [17].

- **Hugging Face:** Utilizado a biblioteca timm para desenvolvimento do modelo e deploy do modelo.
- **PyTorch:** Biblioteca Python para o desenvolvimento da rede neural artificial.
- **Plotly e Matplotlib:** Bibliotecas para visualização de dados

5) **Ferramentas de Processamento de Imagens:** Foram utilizadas as seguintes ferramentas para realizar o tratamento e anotações das imagens de raio-x panorâmica:

- **trainYOLO:** aplicativo web poderoso para gerar anotações em imagens, auxiliando na segmentação semântica.
- **roboflow:** aplicativo web para gerar anotações e aplicar ferramentas de correção comumente utilizadas em processamento de imagens.

B. Processamento da base de dados

A base de dados utilizado neste artigo é um conjunto de 116 imagens anônimas de pacientes, que foram retiradas da *Noor Medical Imaging Center, Qom, Iran* [14]. É um conjunto de imagens de raio-x panorâmicas que abrangem varias situações de saúde bucal, desde pacientes com todos os dentes, até pacientes edêntulos (desprovido de dentes). Esta base de dados foi utilizada em II-F2.

Usando os conceitos de odontologia vistos na seção II-E, foi utilizado a ferramenta trainYOLO para gerar manualmente as anotações dentárias nas imagens de raio-x panorâmica. Utilizando a notação de Palmer conforme a tabela I os dentes de cada imagem foram classificados com valores numéricos utilizando o quadrante Q1-4 e sua posição de 1-8 conforme demonstrado nas “Fig. 10” e “Fig. 11”.

Durante o desenvolvimento das anotações com a ferramenta trainYOLO, foram removidas da base de dados 3 raio-x ficando um total de 113 imagens, pois uma das imagens o paciente possuía ainda seus dentes decíduos, ou seja, era uma imagem com os dentes da fase adulta mais os de leite. Já as outras duas imagens eram de pacientes edêntulos.

Após gerar as anotações, foi necessário tratar as imagens para que elas se adequassem ao modelo de aprendizado de máquina, para isso foi utilizado a ferramenta roboflow no qual

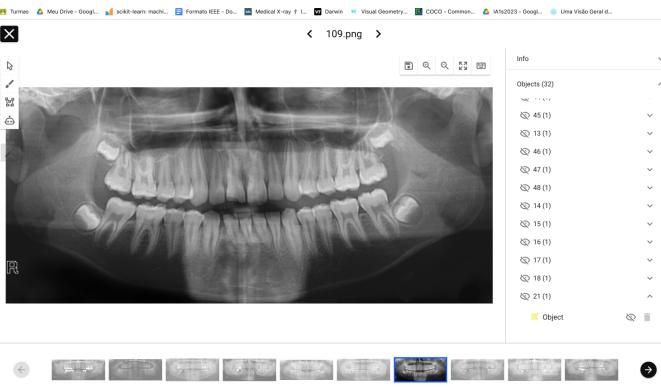


Figura 10. Imagem de raio-x panorâmico número 109 sem anotações.

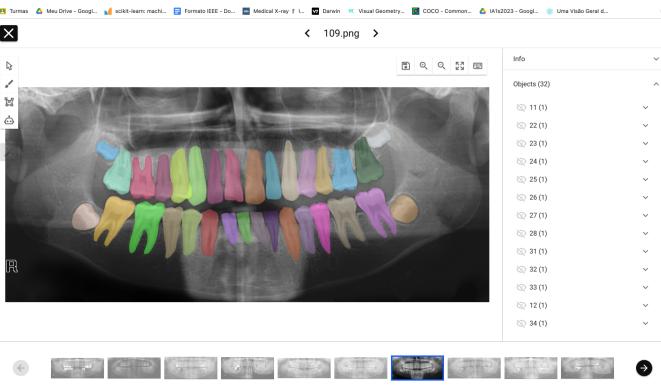


Figura 11. Imagem de raio-x panorâmico número 109 em que possui 32 anotações geradas utilizando a ferramenta trainYOLO.

foi separado as 113 imagens em 2 conjuntos de treino e teste da seguinte forma:

- **treino:** 71% da base de dados, sendo 80 imagens.
- **teste:** 29% da base de dados, sendo 33 imagens.

“Fig. 12” podemos ver como a ferramenta fez a separação dos itens. Neste modelo não será utilizado o terceiro conjunto que é o de validação cruzada por se tratar de um projeto de segmentação semântica.

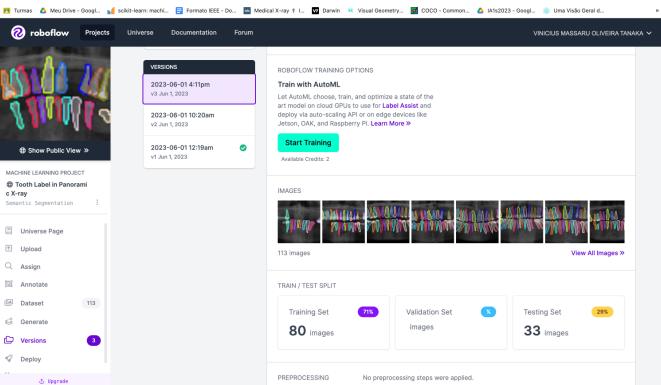


Figura 12. Ferramenta roboflow em que podemos visualizar as imagens que foram separadas em conjunto de teste e treino.

As imagens processadas foram classificadas pixel a pixel conforme a tabela II.

Tabela II
CLASSIFICAÇÃO DOS PIXELS

Valor do Pixel	Classe
0	background
1	11
2	12
3	13
4	14
5	15
6	16
7	17
8	18
9	21
10	22
...	...
16	28
17	31
...	...
24	38
25	41
...	...
32	48

C. Elaboração da Rede Neural Artificial

O pipeline para o desenvolvimento do projeto consiste em 3 partes essenciais em todos os projetos de machine learning:

- **Escolha da Arquitetura:** Decisão de qual arquitetura melhor se adequa ao problema a ser solucionado.
- **Treino:** Treinar o modelo escolhido.
- **Diagnóstico:** Realizar o diagnóstico do modelo para melhorá-lo.

Durante o desenvolvimento da rede neural com base no modelo HRNet IV-A1, houve alguns problemas relacionados às versões das bibliotecas o que conduziu o desenvolvimento do modelo para outra base de rede neural chamada SegFormer IV-A2 [13] [15].

O treino do modelo foi realizado com o framework **PyTorch** juntamente com a biblioteca **timm** do hugging face que contém ferramentas que auxiliam na hora de desenvolver modelos de machine learning.

Com as notações dos pixels já definidas em máscaras das imagens originais, podemos tratar essas informações não estruturadas para se adequar a nossa arquitetura utilizando **datasets** e o **transformers** para treino e validação do modelo.

V. RESULTADOS E DISCUSSÕES

Após realizar o que chamamos de fine-tuning do modelo, foi obtido os seguintes resultados para as arquiteturas **SegFormer** e **UPerNet**:

- **Métrica IoU:** “Fig. 14” e “Fig. 17”, sendo os gráficos das arquiteturas SegFormer e UPerNet, respectivamente.

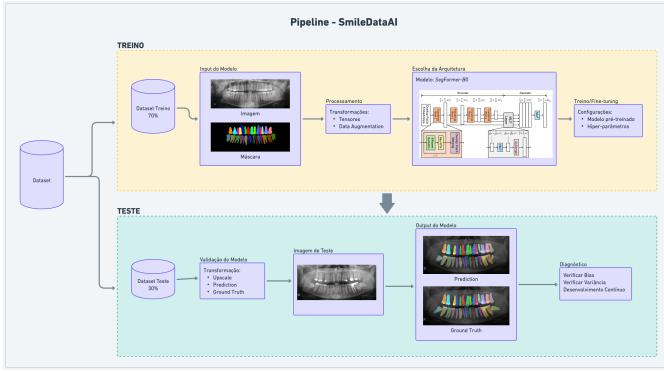


Figura 13. Imagem demonstrando o fluxo de desenvolvimento do projeto, destacando a parte de treino do modelo com fundo amarelo e a parte de teste com fundo verde.

- **Função de perda no conjunto de treino:** “Fig. 15” e “Fig. 18”, sendo os gráficos das arquiteturas SegFormer e UPerNet, respectivamente.
- **Função de perda no conjunto de teste:** “Fig. 16” e “Fig. 19”, sendo os gráficos das arquiteturas SegFormer e UPerNet, respectivamente.

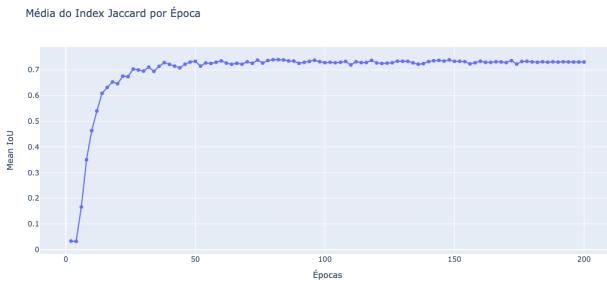


Figura 14. Imagem visualizando o crescimento da média do índice de jaccard durante as épocas para a arquitetura SegFormer “Fig. 8.”

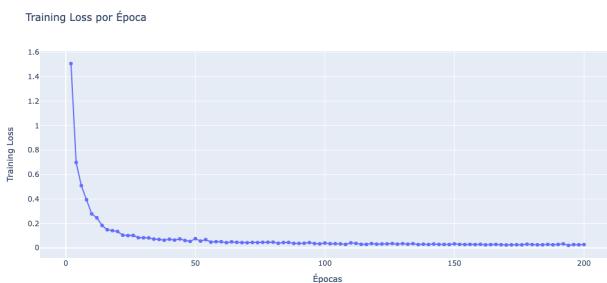


Figura 15. Imagem demonstrando o decaimento da função de perda no conjunto de treino ao longo das épocas para a arquitetura SegFormer “Fig. 8.”

Após obter as estatísticas, podemos realizar um diagnóstico analisando as funções de perda e nossa métrica do índice jaccard. E assim podemos definir os próximos passos para este modelo.

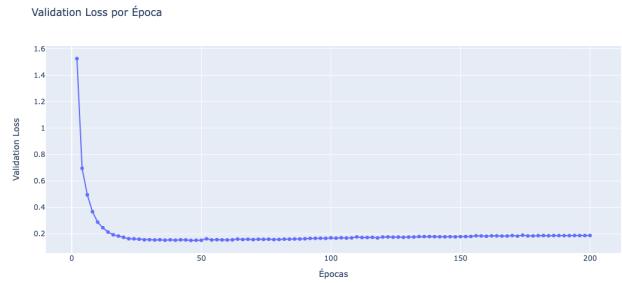


Figura 16. Imagem demonstrando o decaimento da função de perda no conjunto de teste ao longo das épocas para a arquitetura SegFormer “Fig. 8.”

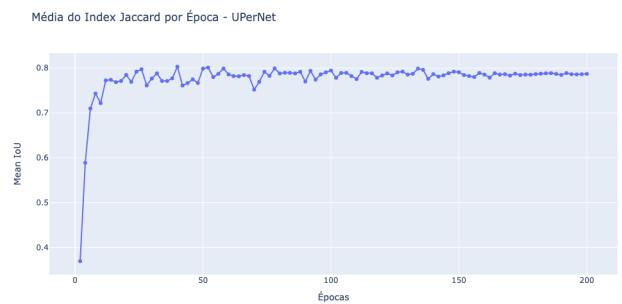


Figura 17. Imagem visualizando o crescimento da média do índice de jaccard durante as épocas para a arquitetura UPerNet “Fig. 9.”

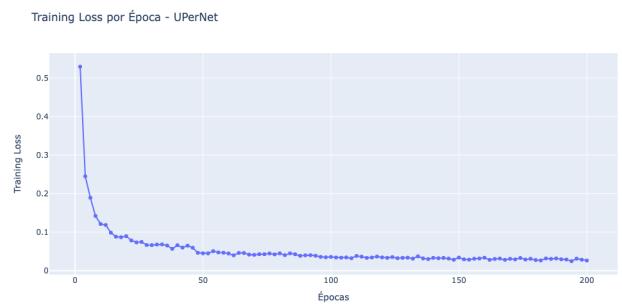


Figura 18. Imagem demonstrando o decaimento da função de perda no conjunto de treino ao longo das épocas para a arquitetura UPerNet “Fig. 9.”

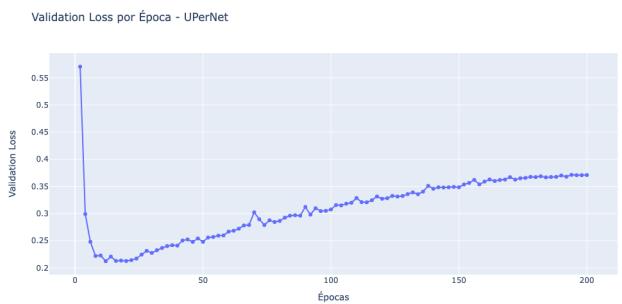


Figura 19. Imagem demonstrando o decaimento da função de perda no conjunto de teste ao longo das épocas para a arquitetura SegFormer “Fig. 9.”

Comparando os gráficos de perda por treino, teste e a média do índice jaccard por épocas das arquiteturas **SegFormer** e **UPerNet**, podemos auferir que o modelo conseguiu se ajustar às imagens de raio-x, porém o modelo SegFormer possui um resultado mais performático no conjunto de validação/teste e que a média do índice de jaccard para ser mais consistente no modelo SegFormer. Vale ressaltar que a quantidade de épocas pode interferir na qualidade do modelo. Nesta versão foi utilizado 200 épocas para o treino, como resultado tivemos um declínio da funções de perda significativa tanto no conjunto de treino quanto no conjunto de teste. Além disso, podemos analisar se nossa arquitetura escolhida possui algum problema de under-fitting ou over-fitting.

Também devemos discutir acerca dos recursos disponíveis para treino do modelo. Neste caso foi utilizado os recursos gratuitos do Google Colab que disponibiliza uma GPU para realizar treino dos modelos.

A quantidade e a qualidade das imagens devem ser levadas em consideração, pois se tratando de imagens de raio-x muitas vezes elas possuem ruídos que podem dificultar o aprendizado da arquitetura.

VI. CONCLUSÕES E TRABALHOS FUTUROS

Com base nas pesquisas e testes realizados, podemos afirmar que é viável a criação do software SmileDataAI para auxiliar os dentistas em diagnósticos personalizados para os pacientes.

Além do auxílio em imagens panorâmicas, é possível ampliar ainda mais as fronteiras que abordar com modelos de visão computacional e inteligência artificial como:

- **Identificar Enfermidades:** Podemos utilizar as mesmas ferramentas aplicadas neste artigo para auxiliar na detecção de lesões, cáries e até mesmo fraturas nas imagens de raio-x. Porém para isso é preciso do auxílio de especialistas nessas doenças.
- **Auditória de Dados Odontológicos:** Muitas clínicas possuem softwares de gerenciamento de clientes, o que abrange um oceano de informações, uma ideia de trabalho futuro seria desenvolver soluções que possam auxiliar os administradores a realizar uma auditória dos dados como, por exemplo, verificar a qualidade do trabalho dos dentistas analisando imagens de raio-x.

Por fim, podemos dizer que a Inteligência Artificial vem transformando a maneira que vemos e interagimos com o mundo de uma maneira positiva. Ela será uma ferramenta importante que poderá alavancar a qualidade e velocidade nos diagnósticos, garantindo uma janela para tomada de decisões mais eficientes.

REFERÊNCIAS

- [1] S. Russell e P. Norving, “Inteligência artificial: Tradução da 3a Edição”, Elsevier Brasil, 2014.
- [2] S. B. Khanagar et al., “Developments, application, and performance of artificial intelligence in dentistry – A systematic review”, Journal of Dental Sciences, vol. 16, no. 1, pp. 1–566, jan. 2021. Disponível em: [Developments, application, and performance of artificial intelligence in dentistry – A systematic review](#). Acessado em: jun. 9, 2023.
- [3] A. C. D. Mazzochi e A. P. P. Virmond, “Inteligência artificial: um conceito futurista no diagnóstico odontológico”, TCC, Centro Universitário Uniguairacá, Guarapuava, Brasil, 2020. Disponível em: [Inteligência artificial: um conceito futurista no diagnóstico odontológico](#). Acessado em: mai 31, 2023.
- [4] Mohammed J. Zaki, Wagner Meira, Jr., “Data Mining and Machine Learning: Fundamental Concepts and Algorithms”, 2nd Edition, Cambridge University Press, March 2020. ISBN: 978-1108473989. Disponível em: [Data Mining and Machine Learning: Fundamental Concepts and Algorithms](#). Acessado em: abr 30, 2023.
- [5] F. A. Faria, “Redes Neurais Artificiais”, Aula ministrada na Universidade Federal de São Paulo, São José dos Campos, São Paulo, Brasil, mar. 29, 2023.
- [6] A. H. Abid, S. Kasaei e M. Mehdizadeh, “Automatic segmentation of mandible in panoramic x-ray”, Journal of Medical Imaging (Bellingham), vol. 2, no. 4, p. oct. 2015. Disponível em: [Automatic segmentation of mandible in panoramic x-ray](#). Acessado em abr. 30, 2023.
- [7] A. Amini, “MIT Introduction to Deep Learning 6.S191: Lecture 3 Convolutional Neural Networks for Computer Vision”, vídeo aula, MIT Deep Learning. Disponível em: [MIT 6.S191: Convolutional Neural Networks](#) Acessado em: mai 21, 2023.
- [8] S. S. Gonçalves, G. L. D. Abreu, S. K. D. Oliveira, “Anatomia Dentária”, Editora UNIFESO, 2020. Disponível em: [Anatomia Dentária](#). Acessado em: abr 21, 2023.
- [9] E. Tiu, “Metrics to Evaluate your Semantic Segmentation Model”, publicado online em: Towards Data Science, ago. 2019. Disponível em: [Metrics to Evaluate your Semantic Segmentation Model](#). Acessado em jun. 1, 2023.
- [10] C. NITR, “MIoU Calculation”, publicado online em: Medium, mai. 9 2020. Disponível em: [MIoU Calculation](#). Acessado em jun. 1, 2023.
- [11] J. E. Cejudo, A. Chaurasia, B. Feldberg, J. Krois e F. Schwendicke, “Classification of Dental Radiographs Using Deep Learning”, Journal of Clinical Medicine, vol. 10, no. 7, abr. 2021. Disponível em: [Classification of Dental Radiographs Using Deep Learning](#). Acessado em abr. 30, 2023.
- [12] W. Park, JK. Huh, JH. Lee, “Automated deep learning for classification of dental implant radiographs using a large multi-center dataset”, Nature, Sci Rep 13, 4862 (2023). Disponível em: [Automated deep learning for classification of dental implant radiographs using a large multi-center dataset](#). Acessado em mai. 30, 2023.
- [13] J. Wang et al., “Deep High-Resolution Representation Learning for Visual Recognition”, in IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 43, no. 10, pp. 3349-3364, out. 1, 2021. Disponível em: [Deep High-Resolution Representation Learning for Visual Recognition](#). Acessado em jun. 13, 2023.
- [14] A. Abdi, S. Kasaei (2020), “Panoramic Dental X-rays With Segmented Mandibles”, Mendeley Data, V2, jul. 01, 2020. Disponível em: [Panoramic Dental X-rays With Segmented Mandibles Dataset](#). Acessado em: abr. 30, 2023.
- [15] E. Xie, W. Wang, Z. Yu, A. Anandkumar, J. M. Alvarez, P. Luo, “SegFormer: Simple and Efficient Design for Semantic Segmentation with Transformers”, University of Hong Kong. Disponível em: [SegFormer: Simple and Efficient Design for Semantic Segmentation with Transformers](#)
- [16] SH. Tsang, “Review — UPerNet: Unified Perceptual Parsing for Scene Understanding”, Medium . Disponível em: ??Review — UPerNet: Unified Perceptual Parsing for Scene Understanding. Acessado em: jul. 8, 2023.
- [17] , “UPerNet, ConvNeXt tiny-sized backbone”, Hugging Face. Disponível em: [UPerNet, ConvNeXt tiny-sized backbone](#). Acessado em: jul. 08, 2023.