



CENTRO DE CIÊNCIAS, TECNOLOGIAS E SAÚDE DO CAMPUS ARARANGUÁ
CURSO DE GRADUAÇÃO EM TECNOLOGIAS DA INFORMAÇÃO E
COMUNICAÇÃO

Lucas Cardoso Benedet

**UM MÉTODO VOLTADO À CLASSIFICAÇÃO DE PADRÕES CONSTRUTIVOS
EM CIDADES POR MEIO DE APRENDIZADO PROFUNDO**

Araranguá
2024

Lucas Cardoso Benedet

**UM MÉTODO VOLTADO À CLASSIFICAÇÃO DE PADRÕES CONSTRUTIVOS
EM CIDADES POR MEIO DE APRENDIZADO PROFUNDO**

Trabalho de Conclusão de Curso submetido ao Curso de Graduação em Tecnologias da Informação e Comunicação do Centro de Ciências, Tecnologia e Saúde do Campus de Araranguá da Universidade Federal de Santa Catarina como requisito parcial para a obtenção do título de Bacharel em Tecnologias da Informação e Comunicação.

Orientador: Prof. Alexandre Leopoldo Gonçalves, Dr.

Araranguá

2024

Benedet, Lucas Cardoso

UM MÉTODO VOLTADO À CLASSIFICAÇÃO DE PADRÕES CONSTRUTIVOS
EM CIDADES POR MEIO DE APRENDIZADO PROFUNDO / Lucas
Cardoso Benedet ; orientador, Alexandre Leopoldo Gonçalves,
2024.

70 p.

Trabalho de Conclusão de Curso (graduação) -
Universidade Federal de Santa Catarina, Campus Araranguá,
Graduação em Tecnologias da Informação e Comunicação,
Araranguá, 2024.

Inclui referências.

1. Tecnologias da Informação e Comunicação. 2. Padrões
construtivos. 3. Vetores densos. 4. Aprendizado profundo.
5. Redes neurais. I. Gonçalves, Alexandre Leopoldo. II.
Universidade Federal de Santa Catarina. Graduação em
Tecnologias da Informação e Comunicação. III. Título.

Lucas Cardoso Benedet

**UM MÉTODO VOLTADO À CLASSIFICAÇÃO DE PADRÕES CONSTRUTIVOS EM
CIDADES POR MEIO DE APRENDIZADO PROFUNDO**

Este Trabalho de Conclusão de Curso foi julgado adequado para obtenção do título de “Bacharel em Tecnologias da Informação e Comunicação” e aprovado em sua forma final pelo Curso de Graduação em Tecnologias da Informação e Comunicação.

Araranguá, 27 de junho de 2024.



Coordenação do Curso

Banca examinadora



Prof Alexandre Leopoldo Gonçalves, Dr.

Orientador



Prof.(a) Marina Carradore Sérgio, Dr.(a)

Universidade Federal de Santa Catarina



Prof Fabrício Herpich, Dr.

Universidade Federal de Santa Catarina

Araranguá, 2024.

Dedico o presente trabalho aos meus pais que me incentivaram e me apoiaram desde sempre a realizar a jornada acadêmica e obter o grau de bacharel.

AGRADECIMENTOS

Tenho a agradecer neste trabalho principalmente meu orientador, Professor Alexandre, o qual sempre esteve presente e disposto a me auxiliar, definindo metas e delimitando os processos para que a realização desta pesquisa ocorresse da melhor forma possível.

Em seguida, agradeço à UFSC, a todo(a)s o(a)s professore(a)s e ao curso de Tecnologias da Informação e Comunicação por toda a base de conhecimento adquirida, o apoio e os conselhos prestados para o desenvolvimento dessa pesquisa.

Agradeço também a minha namorada, Amália, assim como meus pais que também durante todo o processo de desenvolvimento estiveram presentes prestando apoio.

Por fim, agradeço a minha atual empresa e-Topocart, por ceder espaço de trabalho, ideias e elementos essenciais para que este projeto fosse concluído. Assim como, a todo(a)s o(a)s demais que de alguma forma colaboraram com sugestões, ideias ou críticas, para que essa jornada acadêmica fosse finalizada com sucesso.

RESUMO

Os padrões construtivos de casas e prédios dentro de uma cidade desempenham um papel crucial para a gestão urbanística eficiente. Tais padrões influenciam diretamente na arrecadação de impostos, na segurança estrutural e no planejamento urbano, visto que essa classificação é um atributo essencial para o cálculo do valor venal de um imóvel. Com o avanço da tecnologia, novas ferramentas podem ser aplicadas para otimizar esses processos. Sendo assim, este trabalho propõe um método que utiliza redes neurais e aprendizado profundo com o objetivo de identificar e classificar automaticamente imagens das fachadas de imóveis, provendo como resultado final recomendações dos padrões construtivos baseados em uma imagem de fachada de entrada. Para a condução metodológica deste trabalho foi utilizada a *Design Science Research Methodology* (DSRM), tendo como artefato resultante um método. Quanto à avaliação do método proposto, diversos modelos de redes neurais pré-treinados foram testados a partir de um conjunto de dados contendo fachadas e suas respectivas classificações. A partir disso, os resultados demonstraram uma acurácia de mais de 90% quando recomendada pelo menos duas classes. De modo geral, concluiu-se que o artefato gerado neste trabalho é promissor, podendo se tornar um produto com o intuito de melhorar a agilidade e precisão do processo de classificação de padrões construtivos realizado por prefeituras.

Palavras-chave: Padrões construtivos; Vetores densos; Aprendizado profundo; Redes neurais.

ABSTRACT

The building standards of houses and buildings within a city play a crucial role in efficient urban management. These standards directly influence tax collection, structural safety, and urban planning, as this classification is an essential attribute for calculating the taxable value of a property. As technology advances, new tools can be applied to optimize these processes. Therefore, this work proposes a method that uses neural networks and deep learning to automatically identify and classify images of building facades, providing as a final result recommendations for building standards based on an input facade image. For the methodological conduction of this work, the Design Science Research Methodology (DSRM) was used, resulting in a method as the artifact. Regarding the evaluation of the proposed method, several pre-trained neural network models were tested using a dataset containing facades and their respective classifications. As a result, the results demonstrated an accuracy of over 90% when at least two classes were recommended. Overall, it was concluded that the artifact generated in this work is promising and could become a product with the aim of improving the agility and accuracy of the building standards classification process carried out by city halls.

Keywords: Building standard; Embeddings; Deep learning; Neural networks.

LISTA DE FIGURAS

Figura 1 - Classificação de cada padrão construtivo em exemplos de fachadas	21
Figura 2 - Blocos de construção de uma CNN	23
Figura 3 - Erros em (%) de cada modelo de CNN ao longo do tempo na classificação de imagens do conjunto ImageNet	25
Figura 4 - Representação de n <i>embeddings</i> com diferentes contextos	26
Figura 5 - Embedding de imagem por uma CNN	27
Figura 6 - Etapas da classificação supervisionada	28
Figura 7 - Etapas da classificação não supervisionada	29
Figura 8 - Passos executados na identificação dos trabalhos correlatos	30
Figura 9 - Etapas para condução da DSRM	37
Figura 10 - Etapas do método proposto	39
Figura 11 - Representação de um ponto armazenado no <i>Qdrant</i> [®]	41
Figura 12 - Função de pesquisa de vetores	42
Figura 13 - Retorno da função de pesquisa de vetores	43
Figura 14 - Imagem de fachada de classe “Médio Baixo” e representação do <i>embedding</i> correspondente	44
Figura 15 - Representação de todas as classes em 3D	48
Figura 16 - Representação das classes Médio e Médio Baixo em 3D	49
Figura 17 - Representação do restante das classes em 3D	49
Figura 18 - Função de histograma para cada variação de k e n	51
Figura 19 - Foto de fachada com classe “Média” utilizada no primeiro cenário	55
Figura 20 - Foto de fachada com classe “Média” utilizada no segundo cenário	56
Figura 21 - Foto de fachada com classe “Alta” utilizada no terceiro cenário	57

LISTA DE QUADROS

Quadro 1 – Comparação entre os trabalhos correlatos e o presente trabalho	33
Quadro 2 – Abstração de um <i>embedding</i>	41

LISTA DE TABELAS

Tabela 1 - Resultados referentes à Figura 14	44
Tabela 2 - Modelos de rede neural e suas respectivas dimensionalidades	47
Tabela 3 - Histograma para uma imagem de classificação “Média” do conjunto de testes	51
Tabela 4 - Acurácia de todos os modelos para $k = 2$ e $n = 10$	52
Tabela 5 - Matriz de acurácias para cada variação de k e n do modelo DINO-V2-Base	54
Tabela 6 - Ranking de classes retornado para o primeiro cenário	55
Tabela 7 - Ranking de classes retornado para o segundo cenário	56
Tabela 8 - Ranking de classes retornado para o terceiro cenário	57

LISTA DE ABREVIATURAS E SIGLAS

ODS	Objetivos de Desenvolvimento Sustentável
ICT	<i>Information and Communication Technologies</i>
SC	<i>Smart Cities</i>
AI	<i>Artificial Intelligence</i>
LR	<i>Land Registry</i>
PVG	Planta de Valores Genéricos
ANN	<i>Artificial Neural Network</i>
DL	<i>Deep Learning</i>
OMI	Observatório do Mercado Imobiliário
CNN	<i>Convolutional Neural Network</i>
ReLU	<i>Rectified Linear Unit</i>
R-CNN	<i>Region Based Convolutional Neural Networks</i>
UAV	<i>Unmanned Aerial Vehicle</i>
B-CNN	<i>Branch Convolutional Neural Network</i>
DA	<i>Data Augmentation</i>
ML	<i>Machine Learning</i>

SUMÁRIO

1 INTRODUÇÃO.....	14
1.1 OBJETIVOS.....	16
1.1.1 Objetivo Geral.....	16
1.1.2 Objetivo Específicos.....	16
1.2 JUSTIFICATIVA E RELEVÂNCIA DO TEMA.....	16
1.3 ESTRUTURA DO TRABALHO.....	17
2 FUNDAMENTAÇÃO TEÓRICA.....	19
2.1 PADRÕES CONSTRUTIVOS.....	19
2.2 APRENDIZADO PROFUNDO.....	21
2.2.1 Rede Neural Convolucional.....	22
2.3 PROCESSAMENTO DE IMAGENS.....	24
2.3.1 Embeddings de Imagens.....	25
2.3.2 Classificação de Imagens.....	27
2.4 TRABALHOS CORRELATOS.....	29
3 METODOLOGIA.....	35
3.1 CARACTERIZAÇÃO DA PESQUISA.....	35
3.2 DESIGN SCIENCE RESEARCH METHODOLOGY.....	36
4 MÉTODO PROPOSTO.....	39
5 APRESENTAÇÃO DOS RESULTADOS.....	45
5.1 CENÁRIO DE ESTUDO.....	45
5.2 MATERIAIS E MÉTODOS.....	45
5.3 ANÁLISE E DISCUSSÃO DOS RESULTADOS.....	48
5.3.1 Detalhamento do Dataset.....	48
5.3.2 Análises Gerais dos Resultados.....	50
5.3.3 Análises Específicas dos Resultados.....	53
6 CONSIDERAÇÕES FINAIS E TRABALHOS FUTUROS.....	59
6.1 TRABALHOS FUTUROS.....	60
REFERÊNCIAS.....	62
APÊNDICE I.....	67
ANEXO I.....	70

1 INTRODUÇÃO

Os Objetivos de Desenvolvimento Sustentável (ODS) definidos pela Organização das Nações Unidas (ONU) têm como missão acabar com a pobreza, proteger o meio ambiente e o clima, e garantir que todos possam desfrutar de paz e prosperidade. Assim, os ODS são uma ferramenta de grande valor na luta contra a desigualdade social e atuam como guia para o desenvolvimento de políticas públicas que busquem o fortalecimento e prosperidade das cidades como um espaço sustentável (Ospina; Morales, 2020). Seguindo a lógica do ODS número 11, uma cidade sustentável deve garantir assentamentos humanos inclusivos, seguros e resilientes e, para que ocorra isso, faz-se necessário uma governança inteligente que deve ser capaz de lidar e aplicar as Tecnologias da Informação e Comunicações (do inglês *Information and Communication Technologies* - ICTs) emergentes no âmbito de seus governos (Back; Boff, 2022).

Dessa forma, no atual contexto mundial vê-se muito presente o conceito de cidades inteligentes (do inglês *Smart Cities* - SC), ou seja, cidades tecnológicas que oferecem soluções inovadoras para os desafios enfrentados pelas administrações de uma zona urbana por meio das ICTs (Bokhari; Myeong, 2023). Com isso, melhora-se a eficiência operacional dentro das prefeituras e obtém-se informações e dados valiosos com o intuito de fornecer serviço de qualidade à população.

Nesse contexto, os autores Shafiri *et al.* (2024) afirmam que a implementação avançada das ICTs é fundamental para aumentar a eficiência dos recursos, atualizar as infraestruturas, promover o desenvolvimento socioeconômico e, principalmente, integrar o conhecimento humano às tecnologias. Ademais, os autores também complementam dizendo que essas tecnologias oferecem soluções para melhorar aspectos relacionados ao planejamento urbano, projeto, operações e governança, indo de encontro ao ODS 11.

Porém, a implementação de uma cidade inteligente que seja realmente inclusiva e contribua para o desenvolvimento é um desafio, como complementam os autores Back e Boff (2022). A partir disso, como destacado por Freire *et al.* (2021), com a ascensão da Inteligência Artificial (do inglês *Artificial Intelligence* - AI) nos últimos anos, percebe-se variados tipos de aplicações no cenário de cidades inteligentes. Essas aplicações se integram facilmente a diferentes órgãos do governo e constituem uma abordagem personalizável e de grande potencial, pois

viabilizam um gerenciamento mais controlado de setores específicos e respondem de forma mais eficaz às necessidades do cidadão e do município.

Tendo em vista a importância que a tecnologia oferece às cidades, percebe-se a necessidade do desenvolvimento de aplicações ou meios tecnológicos facilitadores, como descrevem Busko e Apolo (2023), a fim de solucionar ou otimizar um trabalho que seja muito sistemático e complexo para funcionários das prefeituras, como o processo de atualização da base de dados cadastrais imobiliários. Um Cadastro Territorial (do inglês *Land Registry* - LR) consiste em uma base de dados única que reúne informações provenientes de diversas fontes, sendo mantida e atualizada por uma ou mais pessoas ou mesmo organizações (Victorino; Amorim; Shimabukuro, 2015). Portanto, integram os dados a serem gerenciados todas as variáveis utilizadas para se calcular o valor de mercado de um imóvel.

Como em um município, a base de cálculo para a arrecadação tributária é o valor do imóvel, e devido à impossibilidade de avaliação individual de cada imóvel em uma cidade, um dos instrumentos utilizados para determinar os valores venais é a avaliação em massa, tecnicamente conhecida como Planta de Valores Genéricos (PGV). A PGV consiste na representação gráfica do perímetro urbano do município, indicando os valores de mercado por metro quadrado dos terrenos. Essa representação é baseada em regiões homogêneas, que são determinadas por uma combinação de atributos (Paiva; Antunes, 2017).

Nesse sentido, um dos atributos que compõem a homogeneidade das regiões é o padrão construtivo. O padrão construtivo é definido a partir de uma análise presencial da qualidade de construção do imóvel, ou seja, torna-se necessário o deslocamento de um funcionário especializado para que sejam registradas fotos das fachadas das edificações de modo que, posteriormente, sejam classificadas por um engenheiro ou fiscal de obras da prefeitura.

A partir disso, por menor que seja a cidade torna-se trabalhoso a classificação do padrão construtivo de inúmeras casas e prédios. Relacionado a isso, inúmeras técnicas podem ser utilizadas para auxiliar na detecção e classificação de imagens, principalmente em soluções baseadas em Redes Neurais Artificiais (do inglês *Artificial Neural Network* - ANN) com características de Aprendizado Profundo (do inglês *Deep Learning* - DL), que facilitem o procedimento de classificação dos padrões construtivos.

Diante deste contexto, a seguinte pergunta de pesquisa foi elaborada: “Como classificar padrões construtivos em cidades considerando técnicas computacionais voltadas ao processamento de imagens?”.

1.1 OBJETIVOS

1.1.1 Objetivo Geral

Em vista disso, este trabalho tem como objetivo geral propor e desenvolver um método voltado à classificação de padrão construtivo de imóveis baseado em aprendizado profundo.

1.1.2 Objetivo Específicos

Para atingir o objetivo geral, os seguintes objetivos específicos destacam-se e são apresentados a seguir:

- 1) Identificar formatos e tipos de classificação de padrões construtivos para coleta de dados;
- 2) Identificar modelos de ANNs capazes de processar imagens de fachada;
- 3) Especificar um modelo de representação e armazenamento de padrões construtivos que permita a execução da tarefa de classificação;
- 4) Elaborar um cenário de estudo que permita avaliar o método proposto neste trabalho.

1.2 JUSTIFICATIVA E RELEVÂNCIA DO TEMA

A gestão eficiente das cidades requer métodos ágeis e precisos para acompanhar o desenvolvimento e crescimento das zonas urbanas. A aplicação de tecnologias emergentes, como ANNs e DL, auxilia no gerenciamento dos dados coletados por uma cidade, visto que esses dados acabam se tornando um desafio de gestão pelo fato de serem obtidos de forma heterogênea e em grande quantidade, comumente chamados de *big data* (Ullah *et al.*, 2024).

Como descrito por Cugurullo *et al.* (2024) as inteligências artificiais são os cérebros das cidades que residem em diversas plataformas digitais e apresentam

soluções para múltiplos domínios urbanos como transportes, segurança, saúde, meio ambiente e urbanismo.

Dessa forma, a obtenção de dados sobre a classificação do padrão construtivo de imóveis é um processo custoso e importante para a arrecadação de impostos das prefeituras. Para tal, faz-se necessário a aplicação de novas tecnologias, oferecendo uma oportunidade para melhorar a eficiência desse processo com o intuito de gerar resultados mais rápidos e precisos.

A otimização do processo de classificação do padrão construtivo contribui diretamente para a prática dos engenheiros responsáveis pela análise do cadastro imobiliário das prefeituras. Ao automatizar a análise por um modelo de AI, tem-se a expectativa de redução significativa do tempo necessário para execução dessa tarefa. Sendo assim, incrementar a precisão e agilidade na classificação, contribui para a implementação eficaz de políticas públicas que promovam o desenvolvimento sustentável e ofereçam maior qualidade de vida para os cidadãos, conforme preconizado no ODS 11.

1.3 ESTRUTURA DO TRABALHO

Este trabalho está organizado em seis capítulos principais, sendo este primeiro responsável por apresentar a introdução do projeto, incluindo a pergunta de pesquisa, os objetivos gerais e específicos a serem alcançados, além da justificativa e relevância do tema.

Na sequência, o Capítulo 2 provê a fundamentação teórica, abordando os conceitos de padrões construtivos, redes neurais artificiais, e técnicas de processamento e classificação de imagens, finalizando com a seção de trabalhos correlatos.

O Capítulo 3 descreve a metodologia da pesquisa adotada como base para a execução deste trabalho.

O Capítulo 4 detalha todas as etapas do método proposto que suportam o processo de classificação de padrões construtivos e cria as bases para a instanciação do mesmo, visando demonstrar seu potencial de aplicação.

Com base nas etapas detalhadas no Capítulo 4, o Capítulo 5 apresenta os resultados obtidos, incluindo exemplos práticos.

Por fim, o Capítulo 6 apresenta as considerações finais, bem como discute as possibilidades de melhorias e direções para trabalhos futuros.

2 FUNDAMENTAÇÃO TEÓRICA

2.1 PADRÕES CONSTRUTIVOS

Um dos grandes desafios dos gestores públicos municipais é garantir uma arrecadação tributária eficiente para que sejam realizadas melhorias urbanas significativas as quais a população precisa e demanda dos governantes. Um município possui diversas formas de arrecadar recursos fiscais, porém a fonte de arrecadação advinda dos impostos sobre os imóveis urbanos tem se mostrado bastante debilitada na maioria das cidades. Conhecer e monitorar o Mercado Imobiliário (MI) local torna-se um grande desafio para as prefeituras por conta da volatilidade das variáveis que interferem no cadastro imobiliário e que definem o valor do imóvel.

A implementação de um Observatório do Mercado Imobiliário (OMI) em uma cidade tem se revelado uma tecnologia fundamental para as prefeituras. Conforme destacam Rodrigues, Borges e Leite (2021), o emprego de geotecnologias na identificação dos elementos urbanos não apenas assegura maior precisão nos dados, mas também resulta na redução dos custos associados a *softwares* de imagens, ao passo que impulsiona o aumento da arrecadação municipal. Esse avanço proporciona uma visão abrangente e em tempo real do mercado imobiliário local, elevando a acurácia nos cálculos, como o valor venal do imóvel, e contribuindo significativamente para o combate à evasão fiscal. Neste contexto, um fator de extrema relevância analisado pelos OMIs na definição do valor venal é o padrão construtivo ou a qualidade da construção do imóvel.

Na construção civil, os padrões dos imóveis são definidos a partir de um conjunto de normas regulamentadas pela Associação Brasileira de Normas Técnicas (ABNT) e estabelecidas pelo Comitê Brasileiro da Construção Civil (CB). Já as normas que definem e orientam a execução de qualquer processo construtivo são definidas pelas Normas Brasileiras (NBR) da Construção Civil (Associação Brasileira De Normas Técnicas, 2005, p. 15). Os padrões arquitetônicos de construção podem diferir de cidade para cidade dependendo de como a prefeitura utiliza as classificações, ou dos fatores considerados na hora de estabelecer o padrão construtivo de uma edificação. Porém, como é descrito na norma NBR 12721, que define um modelo base para especificação, a classificação depende da

tipologia do edifício, ou seja, a construção pode ser do tipo comercial, residencial, uni e multifamiliar. Sendo assim, a norma define a padronização como: Padrão Alto, Padrão Médio, Padrão Baixo e Conjunto Popular. Cada tipo de padrão possui diferentes características as quais são analisadas em um imóvel na hora de enquadrá-lo em algum desses modelos. As características detalhadas de cada tipo de padrão (Figura 1), considerando as tipologias, estão definidas abaixo.

- 1) **Padrão Alto:** São construções de alvenaria, pintadas com uma área construída acima de 150m², possuem esquadrias (mais de uma porta ou janela na fachada frontal), acabamento fino, pé direito alto, condomínios horizontais de alto padrão, e uma diversa geometria retangular. Residência composta de quatro dormitórios, sendo um suíte com banheiro e *closet*, outro com banheiro, banheiro social, sala de estar, sala de jantar e sala íntima, circulação, cozinha, área de serviço completa e varanda e abrigo para automóveis.
- 2) **Padrão Médio:** São construções de alvenaria ou madeira especial, pintadas, com acabamento e pé direito normal e geometria tradicional, uma base retangular e cobertura triangular. A residência normalmente é composta por 3 dormitórios, sendo um suíte com banheiro, banheiro social, sala, circulação, cozinha, área de serviço e varanda, além de abrigo para automóveis.
- 3) **Padrão Baixo:** Normalmente construções de madeira simples ou alvenaria em estado precário, sem pintura ou pintura em estado precário, esquadrias simples (uma porta e uma janela na parte frontal), acabamento inferior ou sem acabamento e geometria tradicional. A residência é composta de dois dormitórios, sala, banheiro, cozinha e área para tanque.
- 4) **Conjunto Popular:** É composto por construções de madeira simples ou alvenaria, área construída abaixo de 80m², condomínio horizontal de baixo padrão (casas iguais ou geminadas). As casas pertencentes ao condomínio normalmente possuem 1 pavimento, 1 dormitório, sala, banheiro e cozinha.

Figura 1 - Classificação de cada padrão construtivo em exemplos de fachadas



Padrão Alto



Padrão Médio



Padrão Baixo



Coinjunto Popular

Fonte: Elaborado pelo autor (2024)

2.2 APRENDIZADO PROFUNDO

A AI é uma área que objetiva permitir que computadores consigam imitar o comportamento humano, sendo capazes de resolver problemas simples e complexos com pouca ou nenhuma intervenção humana (Janiesch; Zschech; Heinrich, 2021). A AI integra várias técnicas e tecnologias que permitem que *software*, máquinas e sistemas percebam, desenvolvam, entendam e aprendam com suas próprias experiências, como informam os autores Ahmed, Jeon e Piccialli (2022).

De modo geral, a AI abrange diferentes ramos e técnicas, entre elas o DL e as ANNs. Conforme Nguyen *et al.* (2019), as ANNs possuem unidades denominadas neurônios, organizadas em camadas, que estão diferenciadas em três categorias: camada de entrada, camadas intermediárias ou ocultas e camada(s) de saída. Os neurônios da camada de entrada são ativados a partir de sensores que captam informações, ou seja, um *input* de dados. Os outros neurônios são ativados através de conexões ponderadas de neurônios de camadas prévias já ativadas (Schmidhuber, 2015).

Como explica Sejnowski (2020), comparando as ANNs com o cérebro humano, há um forte contraste entre os neurônios biológicos e os neurônios dos modelos de AI. Entretanto, muitas características dos neurônios reais são importantes e utilizadas nos modelos de DL. Incluem uma diversidade de células otimizadas para funções específicas como plasticidade sináptica de curto prazo, estados de sono em que o cérebro fica *offline* para se reestruturar e redes de comunicação que controlam o tráfego entre as áreas do cérebro. Tais características podem ser exploradas a fim de criar uma maior sinergia entre o cérebro humano e a máquina, o que pode ser vantajoso tanto para biologia como para engenharia, como informa o autor.

Os modelos de DL vêm sendo prevalentes nos últimos anos com aplicações nas mais variadas áreas. Segundo Lecun, Bengio e Hinton (2015), o DL tem viabilizado grandes avanços na resolução de problemas, resultando em uma abordagem extremamente funcional no descobrimento de estruturas complexas utilizando dados de alta dimensionalidade. Isto o torna aplicável a uma gama de domínios, sejam científicos, empresariais ou governamentais.

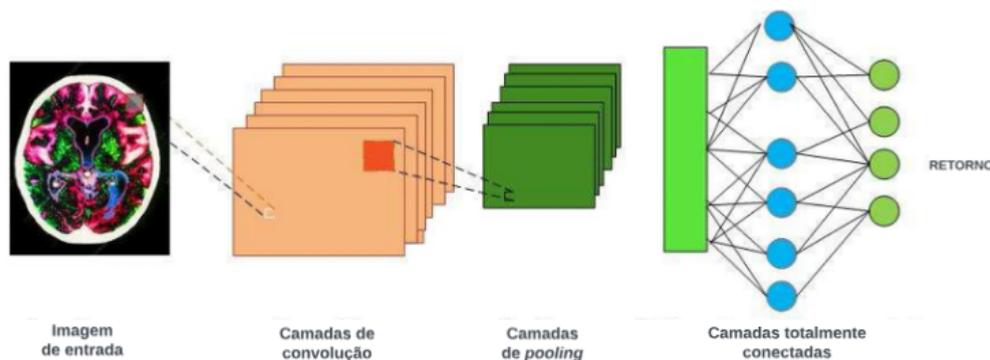
2.2.1 Rede Neural Convolutional

Um papel importante do treinamento supervisionado é a possibilidade de acelerar o processo de aprendizagem de uma rede neural profunda. Como exemplo tem-se a arquitetura de Redes Neurais Convolucionais (do inglês *Convolutional Neural Network* - CNN) que é construída seguindo o modelo de campo visual no cérebro humano. Por conta disso, muitos sistemas de reconhecimento de padrões, imagens e fala obtiveram bons resultados utilizando CNNs (Wang; Fan; Wang, 2021). Seguindo essa ideia, Kattenborn *et al.* (2021) descrevem que o funcionamento de uma CNN pode ser considerado como um mimetismo do córtex animal, onde vários estímulos visuais em diferentes escalas são percebidos pelo campo de visão e as características espaciais presentes nesses estímulos permitem a identificação de objetos.

Sarvamangala e Kulkarni (2022) trabalham a ideia de que CNNs são constituídas de convoluções com pesos e vieses que, semelhantemente aos neurônios, podem aprender. Uma CNN é composta por blocos dos quais os

principais são camadas convolucionais, funções de ativação, *pooling* e camadas totalmente conectadas. Essas camadas estão demonstradas Figura 2 e descritas de forma mais detalhada posteriormente, conforme Sarvamangala e Kulkarni (2022).

Figura 2 - Blocos de construção de uma CNN



Fonte: Adaptado de Sarvamangala e Kulkarni (2022)

Camadas de convolução: Funciona de forma semelhante ao cérebro humano e tem como objetivo extrair características das imagens como bordas, cores, texturas e gradientes. Dessa forma a CNN consegue aprender e distinguir filtros quando encontra essas características. A saída dessa camada é responsável por alimentar as funções de ativação.

Funções de ativação: São responsáveis pela transformação não linear dos dados. Essa camada garante que a representação das características no espaço de entrada serão mapeadas em um espaço de saída diferente. As funções de ativação são as sigmóides, responsáveis por comprimir o valor de uma característica em um intervalo de 0 e 1, as tan hiperbólicas que utilizam um intervalo de -1 e 1 e a unidade linear retificada (do inglês *Rectified Linear Unit* - ReLU) que é responsável por transformar em 0 um valor real negativo.

Camadas de *pooling*: Responsável por reduzir o tamanho espacial dos dados agrupando-os pelo espaço ou tipo de recurso, além de controlar o *overfitting*¹, a translação e a variação rotacional das imagens. O resultado da camada de *pooling* é o particionamento da imagem de entrada, onde cada parte é transformada em um valor único de saída.

¹ Problema que ocorre quando a rede se ajusta excessivamente aos dados de treinamento, apresentando um desempenho significativamente pior em novos dados ou no conjunto de validação/teste.

Camadas totalmente conectadas: Também conhecidas como camadas densas, consistem em um conjunto de nós (neurônios), onde cada nó está conectado a todos os nós da camada anterior. Cada uma dessas conexões possui um peso associado. A saída de cada nó na camada é calculada como a soma ponderada das entradas, onde cada característica de entrada é multiplicada pelo peso correspondente. Em seguida, essa soma passa por uma função de ativação, que introduz não-linearidade ao modelo e permite que a rede aprenda relações mais complexas entre os dados. A função de ativação desempenha um papel crucial no processo de classificação, ajudando a rede a distinguir entre diferentes classes.

2.3 PROCESSAMENTO DE IMAGENS

Neurocientistas acreditam que o sistema sensorial do ser humano consegue se adaptar ao mundo natural a partir de um princípio denominado codificação eficiente. Tratando-se mais especificamente da visão, a codificação eficiente implica em um reconhecimento de uma imagem no cérebro que exige menos processamento e menos energia. Esse princípio deve-se ao fato do ser humano não ser capaz de identificar todos os pixels, mas sim as correlações entre pixels adjacentes os quais criam repetições de segmentos horizontais ou de ângulos, os quais o cérebro é capaz de processar (Sabelli, 2022).

Já no contexto da visão computacional, para reconhecimento de imagens, algumas tarefas são comumente executadas. Entre elas constam a classificação, a segmentação e a detecção de objetos. No caso da classificação, durante a etapa de treinamento, para cada imagem utilizada é também provido um rótulo que a identifica. Desta forma, o rótulo fornecido para dados de entrada visual pode representar uma imagem como um todo, ou partes dessa imagem, como por exemplo, uma janela ou uma porta (Robertson *et al.* 2018).

Diversas arquiteturas de ANNs têm sido propostas para lidar com imagens. Segundo Öztürk, Taşyürek e Türkdamar (2023), as CNNs têm desempenhado papel de relevância devido à facilidade de uso no processamento de imagens, alta capacidade de correspondência e recursos de trabalho que são suportados por GPU. Por conta disso, são muito utilizadas em trabalhos de detecção de objetos e classificação de padrões em imagens. Contudo, a primeira arquitetura de CNN que melhorou significativamente a classificação de imagens a partir de um conjunto

grande de dados, como o ImageNet², foi a arquitetura de rede AlexNet. Com isso, diversos outros modelos surgiram, mas a primeira arquitetura a atingir um resultado de classificação que supera o cérebro humano foi a ResNet, CNN desenvolvida pela Microsoft® em 2015 (Pouyanfar *et al*, 2018). A Figura 3 apresenta a evolução das ANNs quanto à classificação das imagens do ImageNet quanto aos erros percentuais.

Figura 3 - Erros em (%) de cada modelo de CNN ao longo do tempo na classificação de imagens do conjunto ImageNet



Fonte: Pouyanfar *et al*. (2018)

2.3.1 *Embeddings* de Imagens

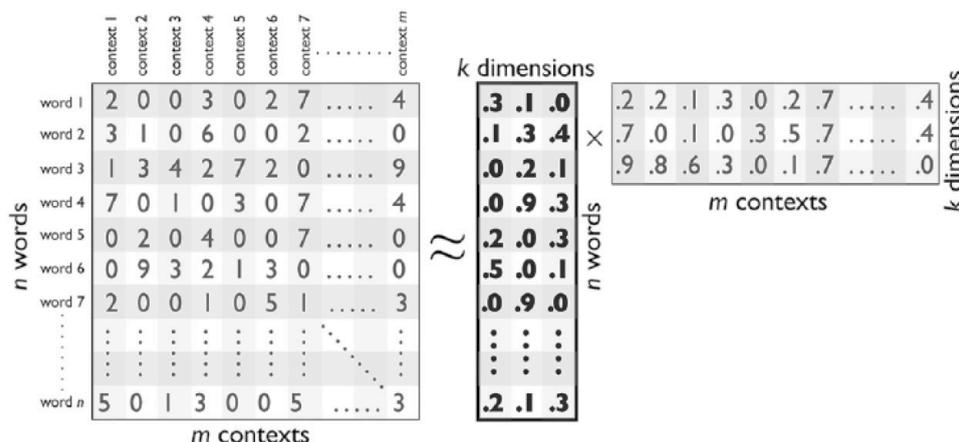
Com o avanço contínuo das ANNs, o DL se estabeleceu como a ferramenta primordial para resolver uma série de problemas, um deles é a similaridade semântica entre as palavras. Conforme apresentado no trabalho de Pennington, Socher e Manning (2014) cada palavra e seu valor semântico podem ser representado em um vetor de valores reais, onde podem ser utilizados em diferentes tipos de aplicações como recuperação de informação, classificação de documentos, resposta a perguntas, reconhecimento de entidade nomeada e análises diversas.

Em um modelo de *embeddings* de palavras, cada palavra é representada como um vetor e são posicionadas no espaço de acordo com a similaridade de seu contexto onde palavras que são similares se encontram mais próximas e palavras com contextos distintos estão a uma distância maior (Kozlowski; Taddy; Evans,

² <https://image-net.org/about.php>

2019). Dessa forma, as representações vetoriais incorporam as palavras em um espaço de recursos n -dimensional, por isso o nome *embedding* e, por conta disso, são capazes de captar os significados das palavras e a relação entre elas e seu contexto (Incitti; Urli; Snidaro, 2023). A partir disso, a Figura 4 a seguir explicita a representação de n *embeddings* de palavras com k dimensões.

Figura 4 - Representação de n *embeddings* com diferentes contextos.



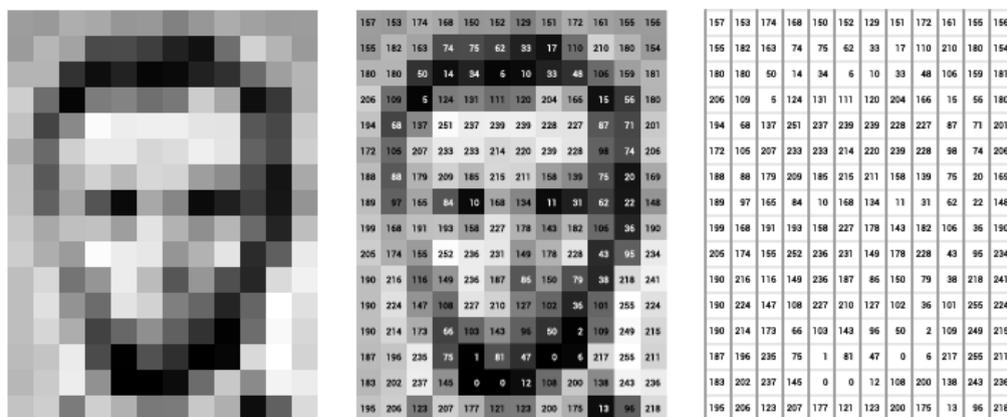
Fonte: Kozlowski, Taddy e Evans (2019)

Este conjunto de definições possibilita a aplicação em diferentes unidades de informação, bastando para isto o provimento de um mecanismo que mapeie determinada entrada para uma representação vetorial densa (*embedding*) n dimensional. Mais especificamente, considerando imagens como uma unidade de informação, as CNNs emergiram como uma arquitetura essencial, especialmente projetada para lidar com dados bidimensionais, como imagens. Assim, o que torna uma CNN tão poderosa é sua capacidade de extrair características relevantes de uma imagem por meio de filtros convolucionais. Esses filtros percorrem a imagem mapeando todos os pixels e com isso capturando padrões e aprendendo representações hierárquicas (Hossain; Sajib, 2019)

Para que ocorra esse mapeamento, a imagem é submetida à uma CNN e as camadas convolucionais são usadas como nível médio para extrair os recursos das imagens (Kiela; Bottou, 2014). Como informam os autores Mandelbaum, Shalev (2016) a extração das características de uma imagem resulta em um *embedding* de grande dimensionalidade, ou seja, cada característica definida pela rede neural é transformada em um valor numérico e armazenada em um vetor. Assim, quanto

mais recursos são extraídos, maior a quantidade de valores dentro do vetor e consequentemente maior sua dimensionalidade. A Figura 5 a seguir demonstra visualmente esse processo e o *embedding* resultante.

Figura 5 - Embedding de imagem por uma CNN



Fonte: Wevers, Smits (2020)

2.3.2 Classificação de Imagens

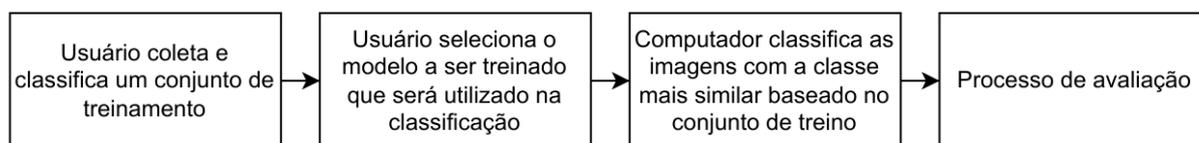
Diariamente são produzidas uma grande quantidade de imagens, e visto que a maior parte dos problemas industriais são problemas de classificação, essas imagens precisam ser classificadas para que possam suportar sistemas nesta área (Nath *et al.*, 2014). A tarefa de classificação de imagens nos últimos anos se tornou um dos campos mais populares da visão computacional devido ao grande sucesso das CNNs com capacidade de extração de características e estrutura flexível para prover soluções para uma ampla variedade de aplicações. Como informam Lei, Pan e Huang (2019), os modelos de classificação de imagens geralmente incluem as arquiteturas Rede Neural de Convolução Completa (FCN), Campo Aleatório de Condição (CRF) e Rede Neural Convolutiva (CNN).

Relacionado à estrutura das CNNs, arquitetura utilizada neste trabalho, Elngar *et al.* (2021) apresentam a classificação de imagens por rede neural como um banco de dados que contém padrões pré-definidos onde quando uma imagem precisa ser classificada ela é apropriadamente rotulada com base nos padrões de imagens armazenadas. Em seguida os autores apresentam diversos tipos de métodos de aprendizagem de imagens por redes neurais, onde as mais conhecidas e utilizadas são a supervisionada e a não supervisionada. Ambos os métodos

necessitam de um conjunto de passos conforme constam nas Figuras 6 e 7. Nesse sentido, uma breve explicação da funcionalidade desses métodos está descrita a seguir.

- **Aprendizagem supervisionada:** Nesta categoria (Figura 6) um conjunto de imagens já classificadas com intervenção humana são utilizadas como conjunto de treinamento para um modelo de rede neural. Dessa forma, quando uma imagem precisa ser classificada, os *pixels* são extraídos dela como representações de superfícies conhecidas que serão tratadas como características. A partir disso, essas características são comparadas com o conjunto das características das imagens classificadas anteriormente para que seja devidamente rotulada.

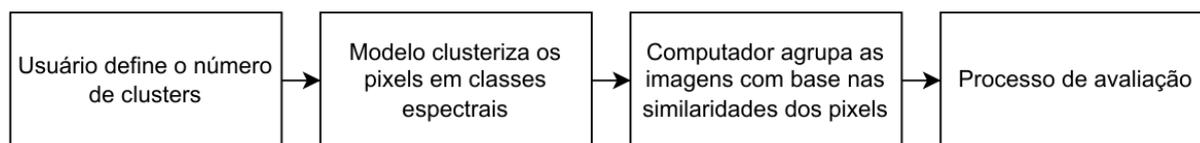
Figura 6: Etapas da aprendizagem supervisionada



Fonte: Adaptado de Elngar *et al.* (2021)

- **Aprendizagem não supervisionada:** Nesta abordagem (Figura 7) está presente o conceito de agrupamento (do inglês *clustering*), ou seja, ocorre a divisão das instâncias de um conjunto de imagens em grupos. Nesta abordagem as imagens não são classificadas anteriormente pois não possuem um rótulo, mas são divididas em grupos por meio de alguma estratégia que analisa as similaridades entre as instâncias, em que as mais próximas são mantidas em um mesmo grupo. O número de grupos pode ser explicitamente definido por um usuário ou por meio de algum parâmetro que define o raio dos agrupamentos, por exemplo. Neste caso, o próprio modelo é capaz de agrupar imagens baseado nas suas propriedades possibilitando, em um segundo momento, o rotulamento dos agrupamentos.

Figura 7: Etapas da aprendizagem não supervisionada



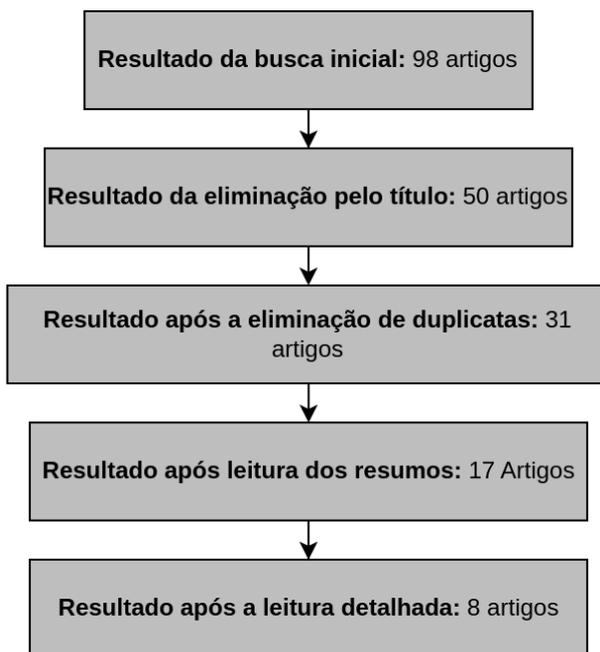
Fonte: Adaptado de Elngar *et al.* (2021)

2.4 TRABALHOS CORRELATOS

Por meio de pesquisas e análises na literatura científica foram obtidos e selecionados trabalhos que se relacionam à classificação ou categorização de imagens, especificamente de fachadas de edifícios ou residências, por meio de Aprendizado de Máquina (do inglês *Machine Learning* - ML) e DL. As pesquisas dos documentos foram realizadas nas seguintes bases científicas: ScienceDirect®, IEEE Xplore®, Web of Science® e SCOPUS®, ACM Digital Library® e no Portal de Periódicos CAPES®. Foram considerados artigos científicos na língua inglesa publicados nos últimos 5 anos completos mais o ano corrente (2018-2023) a partir da seguinte *string* de busca: *facade AND image* AND (building* OR house*) AND (classification OR categorization) AND ("deep learning" OR "Machine learning")*. A Figura 8 a seguir apresenta o resultado dos artigos encontrados nas bases descritas e, em seguida, os trabalhos considerados mais correlatos ao escopo desta pesquisa são resumidos. De modo geral, detalha as etapas utilizadas para seleção dos artigos obtidos por meio da expressão de busca.

A partir de um total de 98 trabalhos, restaram apenas 50 com títulos aderentes à pesquisa. Dos 50 artigos foram descartados aqueles em que os títulos apareciam em mais de uma base de dados, o que resultou em 31 trabalhos. Na sequência, os resumos foram lidos, sendo selecionados 17 trabalhos. Por fim, ao realizar uma leitura mais minuciosa dos 17 artigos restantes, foram selecionados os 8 mais aderentes a este trabalho para efetuar a discussão. O comparativo entre os trabalhos selecionados, além do presente trabalho, estão apresentados no Quadro 1 após o resumo de cada artigo.

Figura 8 - Passos executados na identificação dos trabalhos correlatos



Fonte: Elaborado pelo autor (2023)

O artigo elaborado por Maskeliunas *et al.* (2022) propõe um novo método de classificação de estilos de fachada utilizando a otimização de Pareto aplicada a uma rede de aprendizado profundo baseada em Rede Neural Convolutiva baseada em Região (do inglês *Region Based Convolutional Neural Networks* - R-CNN). O documento tem o objetivo de detectar com precisão os limites das fachadas dos edifícios a partir de imagens de Veículos Aéreos Não-Tripulados (do inglês *Unmanned Aerial Vehicle* - UAV) e classificá-las com base no estilo definido pela taxonomia de fachadas. Os pesquisadores obtiveram um resultado que era dependente das condições climáticas que interferem na qualidade das fotos, sendo 98,41% em configurações de visão clara, 88,11% na chuva e 82,95% quando a imagem estava parcialmente bloqueada por outros objetos.

Já o artigo escrito por Taoufiq, Nagy e Benedek (2020) oferece uma nova abordagem para classificação dos edifícios de uma cidade baseada em fotos das fachadas, a HierarchyNet. O trabalho busca categorizar os edifícios em finalidade funcional e estilos arquitetônicos utilizando uma CNN que pode ser integrada em um *pipeline* de processamento de dados para fornecer informações sobre a funcionalidade e estilo dos edifícios. O modelo HierarchyNet utiliza como base uma CNN Ramificada (do inglês *Branch Convolutional Neural Network* - B-CNN) que se

baseia no paradigma grosseiro para fino, o que melhora a precisão geral para resolução de problemas de processamento de imagens em ambientes urbanos. O modelo foi treinando em diferentes variações de conjuntos de dados e mostrou um desempenho superior em todos eles quando comparado à uma rede B-CNN tradicional.

O trabalho feito por Ottoni *et al.* (2022) propõe uma metodologia de ajuste de hiperparâmetros para Aumento de Dados (do inglês *Data Augmentation - DA*) de classificação de imagens de construção civil que utiliza pequenos conjuntos de dados utilizando DL. Integram as transformações de DA dessa pesquisa Height Shift Range (He), Width Shift Rang (W) and Zoom Range (Z). Os autores obtiveram os resultados a partir da análise do desempenho das redes neurais medidos por modelos de regressão logística. Utilizaram dois estudos de casos: reconhecimento de vegetação nas fachadas e análise de estruturas de coberturas. Para cada um dos estudos de caso foram utilizadas 3 redes neurais da literatura: MobileNet, DenseNet-121 e CNN8. Os melhores resultados foram obtidos, para o modelo CNN8 que atingiu uma média de 92% de acerto para o primeiro estudo de caso, que equivale a 86 imagens classificadas corretamente para um conjunto de dados com 90 imagens.

Bortoloti *et al.* (2022) propõem a construção de um banco de dados de imagens de fachadas de edifícios rotuladas de acordo com seu respectivo uso, obtidas a partir de mosaicos fotográficos originados do Google Street View®. Esse banco de dados foi disponibilizado para um modelo de detecção de objetos baseado em redes de aprendizado profundo chamado de R-CNN. O objetivo do trabalho é expandir o banco de dados para estudo da cidade de Vila Velha com o intuito de mapear a evolução do setor comercial. Cada imagem obtida foi devidamente segmentada e classificada de acordo com seu uso seguindo as opções: residencial, comercial, religioso, público e vago. Com o conjunto de dados pronto, o mesmo foi submetido a duas redes neurais, sendo um modelo treinado do zero e um modelo pré-treinado. Em todos os testes realizados com os modelos em 10 execuções de 40 e 100 épocas o modelo pré-treinado se mostrou pelo menos 20% melhor em ambos os casos, sendo o melhor resultado 86,53% de precisão para um treinamento de 10 execuções em 100 épocas.

O conceito introduzido por Federica *et al.* (2022) busca caracterizar tipos de edifícios de áreas urbanas utilizando dados disponíveis publicamente e imagens

coletadas a partir de trabalho de campo. A pesquisa foi realizada na cidade de Oslo, capital da Noruega, utilizando uma CNN responsável por classificar os edifícios quanto à tipologia de construção para um estudo de avaliação do risco sísmico dessas edificações. Após a obtenção das imagens, as mesmas foram submetidas à uma análise feita por especialistas da área e engenheiros sísmicos, assim definindo-se as tipologias de construção para o parque edificado em Oslo: madeira (T), alvenaria não armada (MUR), betão armado (CR), mista (SRC), aço (S) e outras. Foi utilizado um banco com 5074 imagens das quais 20% foram reservadas para o conjunto de teste e o restante para treinamento do modelo. Com isso, após realizado o trabalho, observou-se que a acurácia do modelo varia de acordo com o material de construção. Todavia, posterior à realização do ajuste fino o modelo obteve uma precisão geral de 82,5% na identificação da classificação das tipologias.

Dini *et al.* (2023) apresentam um estudo baseado no uso de CNN para reconhecimento e classificação de estilo, função, idade e elementos construtivos em edifícios urbanos históricos. Também objetiva a detecção de danos e defeitos nas fachadas causados pelo tempo em uma área apresentada como “*Little China*” na ilha de Java na Indonésia, conhecida por ter muitas referências da cultura chinesa e javanesa em sua arquitetura, porém que sofre com problemas de deterioração. Para isso foram utilizados dois conjuntos de dados, um conjunto primário com dados obtidos diretamente da área histórica de Lasem e um conjunto secundário com dados de imagens do Google Street View[®]. Essas imagens foram submetidas à uma arquitetura VGG-16 juntamente com as classificações utilizadas para as construções, sendo: Estilo Chinês, Estilo Javanês, Estilo Colonial, Estilo Moderno e Estilo Misto. A rede neural foi treinada com 70% do *dataset* e obteve um resultado de 68% de precisão na classificação do caráter visual dos edifícios.

Já Samhour, Al-Arabi e Al-Atrash (2022) apresentam uma nova arquitetura de rede neural responsável pela detecção e classificação de danos em fachadas de construções históricas de pedra amarela baseada no modelo CNN-VGG16, que é capaz de detectar 5 tipos diferentes de classificações nessas estruturas: imagem normal, perda de material, erosão, mudança de cor, e pichação com o objetivo de melhorar a manutenção preditiva e manutenção dessas estruturas. O banco de dados de imagens foi composto a partir de fotos das fachadas dos edifícios obtidas por meio de uma câmera de celular na cidade de Al-Salt na Jordânia, sendo 1024 imagens de fachadas identificadas com problemas arquitetônicos e 216 imagens

sem problemas. Após o modelo treinado e testado as precisões gerais registradas foram de 96% para treinamento e 95% para teste, sendo o valor da perda de 0,1 para ambos os conjuntos.

Por fim, Laupheimer *et al.* (2018) abordam a classificação de fachadas de edifícios baseada na tipologia em um nível alto de abstração, ou seja, classes de uso que são definidas em comercial, híbrido, residencial, uso especial e em construção a partir de DL utilizando CNNs. Para tal, foram utilizados dados reais obtidos do City Survey Office Stuttgart em conjunto com a API do Google Street View® para obtenção das fotos de fachadas. Para o trabalho, as imagens foram submetidas a arquiteturas de redes neurais como VGG19 e ResNet50 que são bastante promissoras em tarefas de classificação. Todas as redes neurais utilizadas para realização dos testes foram pré-treinadas utilizando a base de imagens ImageNet, porém nenhuma delas obtiveram acurácia maior do que 70%, sendo a InceptionV3 o modelo que obteve o melhor desempenho, atingindo 64%. A taxa de erro de 36% de imagens mal classificadas poderia corresponder a recursos mal interpretados pelos modelos, sendo o tipo e quantidade de instâncias de cada classe cruciais para um melhor desempenho do projeto proposto.

Quadro 1 - Comparação entre os trabalhos correlatos e o presente trabalho

Autores	Ano	Método	Conjunto de dados	Métrica	Resultados
Maskeliunas <i>et al.</i>	2022	Pareto optimization, R-CNN	Imagens de fachadas a partir de UAVs	Acurácia	Melhor resultado foi de 98,41%
Taufiq, Nagy e Benedek	2020	HierarchyNet.	Fachadas de edifícios com respectivas finalidades funcionais e estilos arquitetônicos	Acurácia e perda	HierarchyNet se saiu melhor em comparação com os outros modelos em todos os conjuntos de dados, obtendo mais de 80% de precisão
Otoni <i>et al.</i>	2022	Data augmentation para MobileNet, DenseNet-121 e CNN8	Imagens de construção civil	Acurácia	Melhor resultado de 92% para CNN8
Bortoloti <i>et al.</i>	2022	R-CNN	Imagens de fachadas de edifícios rotuladas de acordo com seu respectivo uso	Acurácia	86,53% de precisão para um treinamento de 10 execuções em 100 épocas.
Federica <i>et al.</i>	2022	CNN	Imagens de edifícios de áreas urbanas	Acurácia	82,5% na identificação da classificação das tipologias.
Dini <i>et al.</i>	2023	CNN	Fotos de edifícios urbanos históricos	Acurácia	68% de precisão na classificação do caráter visual dos edifícios.
Samhuri, Al-Arabiati e Al-Atrash	2022	CNN-VGG16,	Fotos de danos em fachadas de construções históricas de pedra amarela	Acurácia	96% para treinamento e 95% para teste
Laupheimer <i>et al.</i>	2018	VGG19, ResNet50 e InceptionV3	Fotos de fachadas de edifícios obtidas a partir do City Survey Office Stuttgart em conjunto com a API do Google Street View®	Acurácia	InceptionV3 o modelo que obteve o melhor desempenho, atingindo 64%
Este trabalho	2024	Diversos modelos, como Dino, Resnet, Vgg, Alexnet, Embeddings e banco de dados vetorial	Fotos de fachadas de imóveis obtidas a partir da prefeitura de Criciúma	Acurácia	Resultados de mais de 90% para $k=2$ em quase todos os modelos

Fonte: Elaborado pelo autor (2024)

Para fins de comparação, todos os demais trabalhos correlatos apresentaram métodos parecidos, utilizando como base CNNs. Além disso, todos apresentaram a utilização de um banco de dados previamente constituído para utilização de treinamento de redes neurais. Com o intuito de aferir a qualidade dos resultados, todos utilizaram métricas de acurácia e, no geral, apresentaram resultados positivos, sendo que apenas dois deles não atingiram pelo menos 70% de acurácia.

Contudo, as técnicas utilizadas para o desenvolvimento deste trabalho apresentam um diferencial quando comparado aos demais. Para o desenvolvimento desta pesquisa optou-se pelo uso do conceito de *embeddings* e banco de dados vetorial. Isto implica diretamente no fato de que novas imagens podem ser facilmente adicionadas à base de treinamento (entendida como uma base de conhecimento). Ademais, este trabalho permite que qualquer modelo pré-treinado capaz de gerar *embeddings* de imagens seja utilizado nas etapas de treinamento e testes, demonstrando a flexibilidade da proposta.

3 METODOLOGIA

3.1 CARACTERIZAÇÃO DA PESQUISA

Para Gerhardt e Silveira (2009) uma pesquisa só se inicia quando existe alguma dúvida ou problemática a ser respondida e, para a realização dessa pesquisa, é necessário uma metodologia, ou seja, um caminho a ser percorrido para que se obtenha um resultado esperado ou não. Nesse sentido, o presente trabalho se utiliza de uma metodologia voltada à pesquisa aplicada, tecnológica e bibliográfica desenvolvida com a finalidade de viabilizar a classificação de padrões construtivos.

Este trabalho caracteriza-se como pesquisa aplicada, tendo em vista que tal pesquisa normalmente se origina a partir de um problema ou dificuldade que precisa ser solucionada, sendo vista como forma de promover inovação e desenvolvimento científico para que seja transformada em conhecimentos capazes de serem aplicados em casos reais (Vieira; Leite; Kuhn, 2023).

Nesse contexto, este trabalho não apenas se alinha a essa abordagem, mas também apresenta a intersecção entre pesquisa aplicada e tecnológica, buscando solucionar um problema real de modo a contribuir com o avanço tecnológico. Tangenciando este fato, Júnior (2014) explica que o objetivo da pesquisa tecnológica é a produção e desenvolvimento de algum artefato com foco no controle da realidade. Afirma também que, invariavelmente, esse procedimento tem como produto final o desenvolvimento de uma nova tecnologia.

Ademais, também foi elaborada uma pesquisa bibliográfica a respeito do tema proposto. Como descrito por Lima e Mito (2007), a pesquisa bibliográfica é realizada com o intuito de fundamentar teoricamente o objeto de estudo, fornecendo elementos e informações que subsidiam a análise dos dados e resultados obtidos. O mapeamento das obras fundamenta-se nos critérios elaborados a partir da revisão da literatura, do referencial teórico, da investigação e do problema apresentado como fio condutor da pesquisa (Graziotin; Klaus; Pereira, 2020). Dessa forma, realizou-se um conjunto ordenado de procedimentos para coleta do material bibliográfico analisado, sendo estes descritos na Seção 2.4 que culminaram com a apresentação dos trabalhos correlatos, mas também geraram insumos para a elaboração dos Capítulos 1 e 2.

Por fim, um trabalho científico pressupõe a utilização de uma metodologia conferindo rigor na execução das etapas a serem executadas, assim como qualidade e capacidade de explicação dos resultados obtidos. A próxima seção detalha a metodologia utilizada neste trabalho.

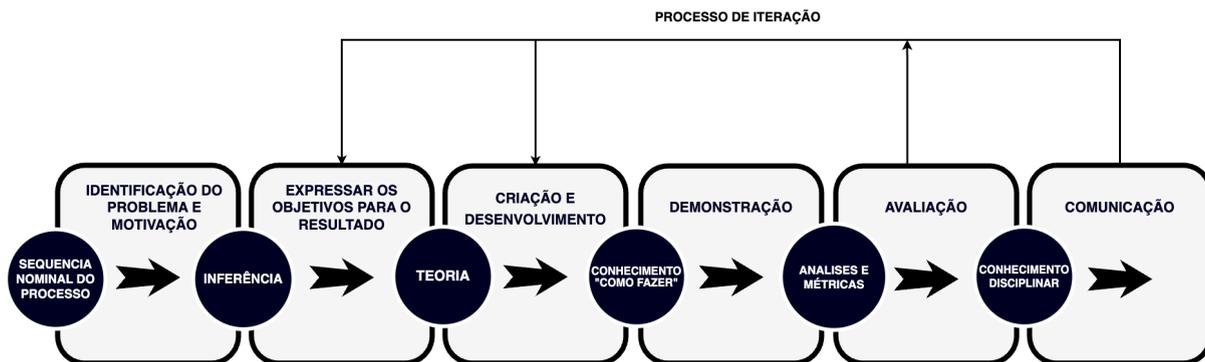
3.2 DESIGN SCIENCE RESEARCH METHODOLOGY

A metodologia proposta para o desenvolvimento deste trabalho fundamenta-se na abordagem da *Design Science* (DS). Como declaram Engström *et al.* (2020), a DS é um paradigma de pesquisa ou base epistemológica, usada comumente em campos de sistemas de informações e áreas da engenharia. Os autores também afirmam que este paradigma busca resolver problemas do mundo real, auxiliando profissionais a desenvolverem soluções para problemas de domínios específicos.

Como explicado por Hevner e Chatterjee (2010), a partir da DS busca-se criar inovações estabelecidas a partir de ideias, práticas, capacidades técnicas e produtos necessários para realizar a análise, o design, a implementação e o uso de sistemas de informação de maneira eficiente e eficaz. Ademais, a DS como forma de aquisição de conhecimento, impacta e é impactada pelas decisões de *design* tomadas em relação à metodologia de desenvolvimento do sistema utilizado e às capacidades funcionais, conteúdos de informação e interfaces humanas implementadas no sistema de informação.

Por consequência, mais especificamente, no que tange a metodologia, tendo como base a DS, para este trabalho será utilizada a *Design Science Research Methodology* (DSRM). Esta metodologia é composta por um fluxo de execução com 6 etapas bem definidas (Peffer *et al.*, 2007). Todavia, como destacado por Abdurrahman e Mulyana (2020), não há um ponto de entrada específico reconhecido nos processos da DSRM, ou seja, não necessariamente o projeto deve iniciar na etapa 1 e terminar na etapa 6. Além disso, ainda sugerem que o procedimento pode ser realizado de forma diferente ao depender do problema e do objetivo final. A seguir são destacadas as etapas do fluxo de execução da DSRM (Figura 9) e, posteriormente, uma descrição explicativa de como cada etapa se encaixa neste trabalho.

Figura 9 - Etapas para condução da DSRM



Fonte: Adaptado de Peffers *et al.* (2007)

1. **Identificação do problema e motivação:** Descrição, explicitação do problema de pesquisa e racionalização do valor da solução proposta. A descrição do problema é utilizada para criação de um artefato. Neste caso, a identificação do problema ocorreu a partir de uma pesquisa bibliográfica que possibilitou criar as bases para a elaboração dos Capítulos 1 e 2, assim como a elaboração da pergunta de pesquisa. O aprimoramento da pesquisa, mais direcionada à solução do método proposto, resultou na Seção 2.4 de trabalhos correlatos. Sendo possível observar um *déficit* de trabalhos relacionados ao uso de tecnologias para classificação automatizada de padrões construtivos.
2. **Expressar os objetivos para o resultado:** Refere-se à conclusão das ideias e do resultado obtido a partir da problemática identificada, bem como da compreensão do que é provável e praticável durante a pesquisa. Os objetivos deste trabalho foram delineados com base em uma análise dos trabalhos obtidos a partir da revisão bibliográfica e, também, a partir de uma análise mais específica visando identificar os trabalhos correlatos. Estes forneceram evidências importantes para orientar o caminho a ser seguido e identificar as potenciais contribuições que este trabalho poderia oferecer. Tal abordagem fundamentada em trabalhos anteriores permitiu uma compreensão mais ampla do campo de estudo e auxiliou na identificação de lacunas de pesquisa a serem preenchidas, além de fornecer um contexto sólido para a proposição de soluções inovadoras e eficazes.

3. **Criação e desenvolvimento:** Refere-se à construção do artefato, podendo ser um modelo, um método, um *framework*, uma instanciação, entre outros. Este procedimento consiste em definir a função do artefato e seu *design* e, posteriormente, projetar o artefato real. Assim, o resultado final deste projeto que materializa ou instância o método proposto, é representado por uma ferramenta de recomendação, nível de protótipo, que permite sugerir a relevância dos padrões construtivos visando tomadas de decisões mais rápidas e assertivas por parte do usuário ou de um sistema automatizado.
4. **Demonstração:** Demonstração prática do artefato para resolução de uma ou mais instâncias do problema. Esta demonstração foi realizada a partir de um conjunto de dados utilizados para testes em que cenários de estudo foram propostos, sendo detalhados na Seção 5.1 e analisados na Seção 5.3.
5. **Avaliação:** Esta etapa consiste na avaliação e quantificação do quão bem o artefato contribui na promoção do resultado ao problema. Por exemplo, no cenário de AI e ML, a avaliação é realizada comparando o resultado real obtido através da utilização do artefato na fase de demonstração com os objetivos finais esperados. No caso deste trabalho, a avaliação foi realizada por meio da equação de acurácia que considera a porcentagem de acerto do modelo de classificação implementado no método proposto, onde TP representa o total de resultados positivos e TN o total de resultados negativos conforme a Equação 1:

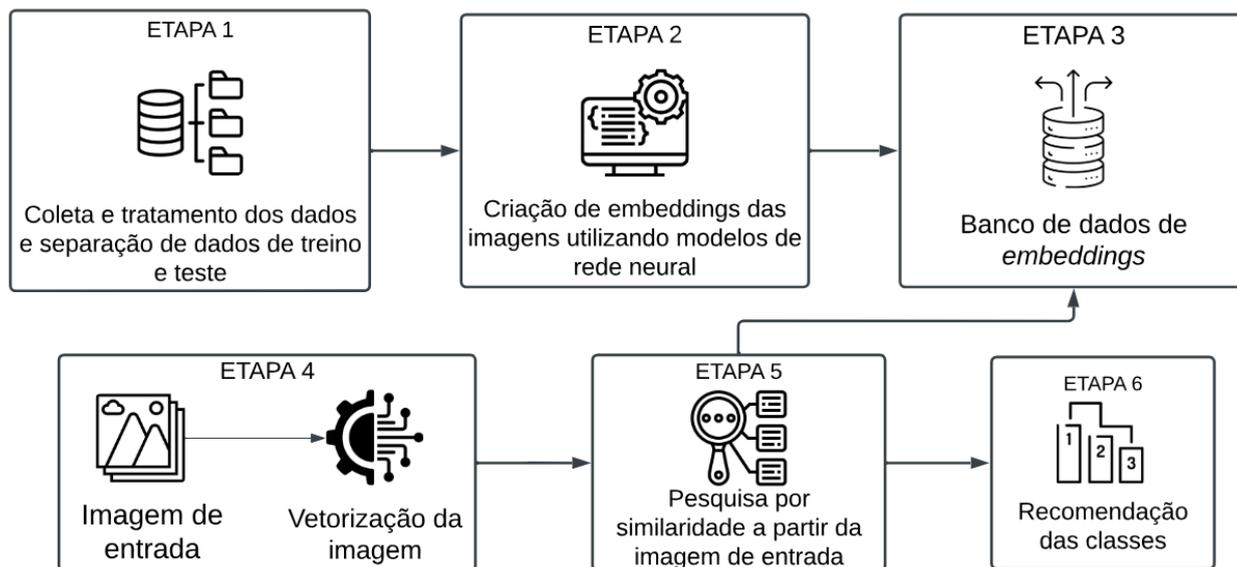
$$Acurácia = \frac{TP}{TP + TN} \quad (1)$$

6. **Comunicação:** Esta etapa representa a comunicação do problema e do seu significado, o artefato construído, o rigor do esquema modelado, e sua eficácia e inovação. A comunicação do trabalho realizado, para além deste próprio documento, pretende ser efetuada por meio da divulgação e publicação do trabalho como artigo científico em periódicos ou conferências de tecnologia.

4 MÉTODO PROPOSTO

O método proposto neste trabalho consiste na realização de seis etapas, sendo detalhadas e diagramadas conforme a Figura 10. A primeira etapa consiste na criação de uma base de dados. Na segunda etapa são criados vetores densos (*embeddings*) a partir das imagens coletadas de fachadas. A seguir, na terceira etapa, os *embeddings* são armazenados em um banco de dados vetorial. Na quarta etapa, após a estruturação do banco, uma imagem de entrada é transformada em *embedding* para permitir a busca no banco de dados por imagens similares. Na quinta etapa, ocorre a consulta ao banco de dados vetorial em que, a partir de uma imagem de entrada, são retornadas as imagens similares. Para a geração dos *embeddings* serão utilizados modelos para processamento de imagens que serão discutidos adiante. E, por fim, na sexta etapa, considerando o resultado da consulta, são avaliadas a(s) classe(s) mais frequentes permitindo recomendar uma ou mais classes para determinada imagem de entrada.

Figura 10 - Etapas do método proposto³



Fonte: Elaborado pelo autor (2024)

³ As imagens dos ícones foram obtidas a partir do site <https://www.flaticon.com/br/>

ETAPA 1: Elaboração dos Conjuntos de Dados

A primeira etapa consiste na coleta de dados para criação de uma base voltada à classificação de fachadas de acordo com o conceito de padrão construtivo. Para isso, nesta pesquisa foram disponibilizados dados da prefeitura de Criciúma, através da elaboração de uma declaração assinada (Anexo I) pelo próprio município, permitindo a utilização dos dados contendo fotos de fachadas diversas com a respectiva classificação. Vale ressaltar que as classificações contidas no banco de dados correspondem àquelas definidas pela própria prefeitura de Criciúma, e não seguem a padronização do modelo de base definido pela NBR. Dessa forma, a prefeitura considera cinco classes, sendo estas: “Alto”, “Médio Alto”, “Médio”, “Médio Baixo” e “Baixo”. De modo geral, o padrão construtivo é determinado por engenheiros especialistas em avaliações de imóveis, sendo descrito mais detalhadamente na Seção 2.1. A partir das imagens coletadas estas são divididas em dois conjuntos de dados. O primeiro será armazenado no banco de dados para servir tanto à etapa de treinamento como uma base de conhecimento para futuras aplicações. Já o segundo, refere-se ao conjunto destinado à fase de testes, visando avaliar a acurácia do método proposto.

ETAPAS 2 e 3: Vetorização das Imagens Coletadas e Criação de um Banco de Dados Vetorial

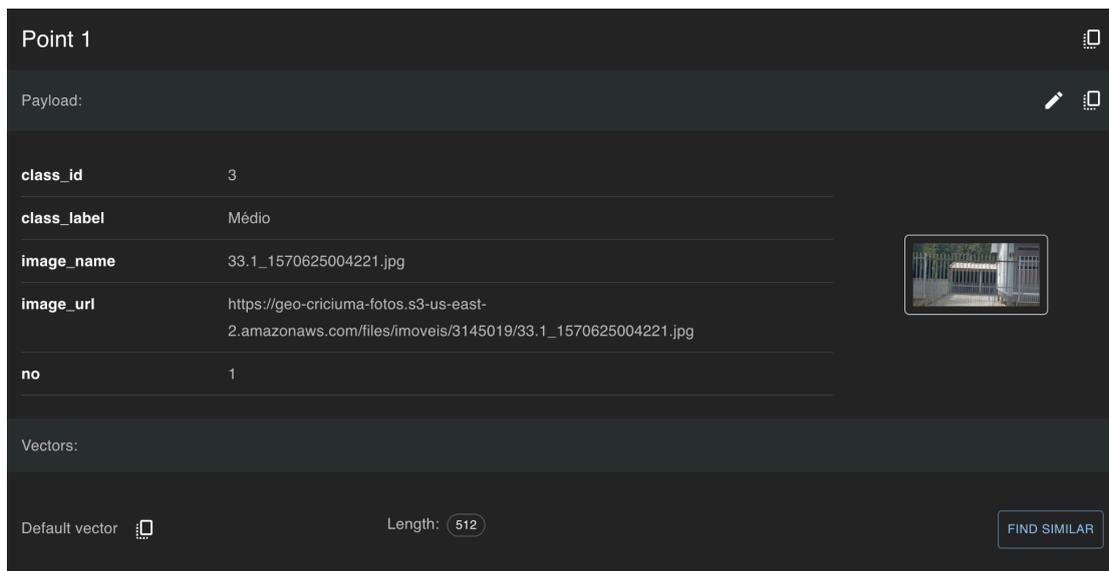
A seguir, para as imagens pertencentes ao conjunto de treinamento são gerados respectivos *embeddings*. O Quadro 2 apresenta um exemplo de um *embedding* indicando as primeiras cinco posições e as cinco últimas posições. Vale mencionar que o tamanho do *embedding* é determinado pelo modelo de processamento de imagem suportado por determinada arquitetura de rede neural. Os *embeddings* criados são então armazenados no Qdrant®, um banco de dados vetorial desenvolvido com tecnologia de pesquisa de similaridade vetorial avançada e de alto desempenho, capaz de suportar vetores de diferentes dimensionalidades.

Quadro 2 - Abstração de um *embedding*

-0.00027024176	0.050940868	0.0056790127	0.024202112	-0.0017601439
...				
0.050506514	-0.058323905	0.0047850953	-0.0037593462	-0.053328477

Fonte: Elaborado pelo autor (2024)

No *Qdrant*[®], o armazenamento ocorre em uma *collection*, ou seja, um espaço para armazenamento de pontos, onde cada ponto representa tanto o *embedding*, quanto outros dados relacionados, por exemplo, o nome, a *url* e a classe da imagem. A Figura 11 mostra a representação de um ponto armazenado em uma *collection*.

Figura 11 - Representação de um ponto armazenado no *Qdrant*[®]

Fonte: Elaborado pelo autor (2024)

ETAPA 4: Vetorização de Imagens de Entrada

A quarta etapa possui duas funções essenciais, sendo a primeira voltada ao teste do método proposto e a segunda a utilização em fluxo regular. Todavia, ambos têm o mesmo fluxo, ou seja, considerando uma imagem de entrada esta é transformada em um *embedding* que será utilizado nas etapas seguintes.

Vale ressaltar que o teste executa esta etapa n vezes, determinada pelo número total de imagens que constam no conjunto de teste. Ao final, a partir de

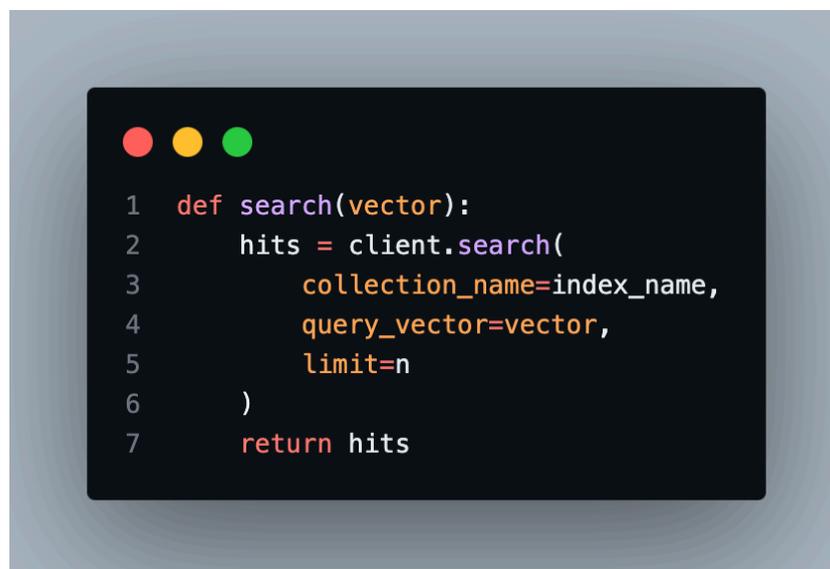
cada imagem, atinge-se a acurácia final permitindo a avaliação do comportamento geral do método.

Com um banco de dados otimizado e preparado, foi desenvolvido um *script* de *searching*, responsável por receber uma imagem de entrada, que será solicitada a partir de uma demanda de trabalho dos usuários. Esta etapa é necessária para possibilitar que os engenheiros responsáveis consigam enviar diretamente para a aplicação uma imagem de fachada. Desta forma, a imagem de entrada é transformada em um *embedding* seguindo a mesma lógica da Etapa 2.

ETAPA 5: Pesquisa por Similaridade

Nesta etapa, o *embedding* gerado a partir da imagem de entrada descrito na Etapa 4 é submetido à uma função responsável por realizar uma pesquisa por similaridade no banco de dados vetorial. A função é representada na Figura 12.

Figura 12 - Função de pesquisa de vetores

A screenshot of a code editor with a dark background and light-colored text. The code is a Python function named 'search' that takes a 'vector' as input. It uses a 'client' object to search a specific 'collection_name' (index_name) for a 'query_vector' (vector) with a 'limit' of 'n'. The function returns the search results, 'hits'.

```
1 def search(vector):
2     hits = client.search(
3         collection_name=index_name,
4         query_vector=vector,
5         limit=n
6     )
7     return hits
```

Fonte: Elaborado pelo autor (2024)

A função acima recebe como entrada um *embedding* para então realizar uma pesquisa baseada em similaridade por meio do método *search* disponibilizado pelo banco de dados vetorial. Na sequência, o banco de dados retorna as informações solicitadas de acordo com um limite n estabelecido, em que n indica os vetores mais similares ao vetor de entrada. O retorno da função para $n=1$ é representado a seguir

(Figura 13). Percebe-se na imagem, por exemplo, o rótulo e o número de identificação da classe que, neste caso, refere-se a classe de padrão construtivo “Médio”.

Figura 13 - Retorno da função de pesquisa de vetores



```

1 [
2   {
3     "ScoredPoint": {
4       "id": 1,
5       "version": 0,
6       "score": 1.0,
7       "payload": {
8         "class_id": 3,
9         "class_label": "Médio",
10        "image_name": "33.1_1570625004221.jpg",
11        "image_url": "https://geo-criciuma-fotos.s3-us-east-2.amazonaws.com/files/imoveis/3145019/33.1_1570625004221.jpg",
12        "no": 1
13      },
14      "vector": null,
15      "shard_key": null
16    }
17  }
18 ]

```

Fonte: Elaborado pelo autor (2024)

ETAPA 6: Ranqueamento das Classes

Para realizar o ranqueamento das classes, é essencial analisar a lista de *embeddings* obtidos na Etapa 5. Nela, são identificadas as classes mais frequentemente associadas, considerando os n vetores mais similares. O parâmetro n deve ser definido pela aplicação ou usuário que realizará a busca. Em geral, a utilização de $n=100$ se mostra adequado, visto que, de modo geral, as primeiras imagens (*embeddings*) recuperadas a partir de um banco de dados vetoriais são as mais relevantes. Essa contagem é fundamental para determinar a porcentagem de adequação de cada classe à imagem de entrada, organizando-as da mais alta para a mais baixa. A quantidade de classes que podem ser retornadas é flexível e depende do parâmetro k (número de classes), determinado exclusivamente pela função de retorno de vetores similares do banco de dados.

De forma a apresentar o funcionamento do sistema de ranqueamento, a Figura 14 ilustra uma imagem de entrada que possui a classificação “Médio Baixo” e uma abstração do *embedding* correspondente a ela. Posteriormente, a Tabela 1

exemplifica o resultado da aplicação, mostrando as recomendações com base na frequência das classes, organizadas de forma decrescente.

Figura 14 - Imagem de fachada de classe “Médio Baixo” e representação do *embedding* correspondente



-3.33952419e-02	-1.33203268e-02	-2.28146203e-02	4.10108864e-02	-1.72285810e-02
...				
2.68088095e-02	-1.84104517e-02	4.69805673e-02	4.21059877e-02	-3.95085961e-02

Fonte: Elaborado pelo autor (2024)

Tabela 1 - Resultados referentes à Figura 14

Classe	Quantidade
4 - Médio Baixo	83
3 - Médio	14
5 - Baixo	2
2 - Médio Alto	1

Fonte: Elaborado pelo autor (2024)

Para o caso acima, foram retornados os 100 *embeddings* que mais se aproximavam ao *embedding* que representa a imagem da Figura 14. Com base nos retornos obtidos, 83 casos representaram a classe “Médio Baixo” e, portanto, essa seria a classe mais recomendada pelo sistema.

5 APRESENTAÇÃO DOS RESULTADOS

5.1 CENÁRIO DE ESTUDO

Este trabalho baseia-se na aplicação de modelos pré-treinados de CNNs capazes de realizar classificações de imagens de padrão construtivo. O padrão construtivo é variável e o modo como é definido depende dos métodos de trabalho da prefeitura responsável por classificá-lo. Para este projeto, como os dados disponibilizados foram da prefeitura da cidade de Criciúma, utilizou-se, também, o modelo de padrão construtivo no qual o município trabalha.

O cenário de estudo consiste em analisar e implementar diferentes tipos de modelos de ANNs pré-treinados para realizar processamento de imagens a fim de classificar automaticamente e recomendar as classes de padrões construtivos a partir de imagens de interesse. Tem como objetivo auxiliar e agilizar o trabalho de profissionais qualificados que sejam designados para realizar tais classificações de forma manual.

Para conduzir os experimentos propostos, foram exploradas diferentes abordagens de classificação, incluindo métodos de similaridade vetorial e resultados derivados do próprio modelo treinado. Além disso, uma variedade de redes neurais foi testada para determinar qual delas se adequaria melhor ao estudo em questão, visando alcançar classificações mais precisas.

Todos os materiais e métodos empregados na obtenção dos resultados, juntamente com as técnicas exploradas, estão detalhados na seção de materiais e métodos que será apresentada a seguir neste trabalho. Essa seção oferecerá uma visão aprofundada das estratégias adotadas, fornecendo um contexto abrangente para os resultados discutidos posteriormente.

5.2 MATERIAIS E MÉTODOS

Para a realização deste trabalho foram utilizados dados dispostos pela prefeitura de Criciúma que continham fotos de fachadas de casas, prédios e estabelecimentos diversos, além da respectiva classificação do padrão construtivo de cada imagem. No conjunto de dados disponibilizado foram realizados

tratamentos simples retirando as imagens as quais não puderam ser baixadas ou visualizadas, resultando, assim, em um *dataset* com 1421 imagens, sendo que a quantidade de imagens para cada classificação foi de 763 imagens com a classificação “Médio Baixo”, 454 com a classe “Médio”, 98 para “Médio Alto”, 78 para “Baixo” e as 28 imagens restantes para a classe “Alto”. Vale ressaltar que estes quantitativos fazem parte do *dataset* obtido junto à Prefeitura de Criciúma estando, como pode ser observado, desbalanceado.

Com a estruturação do *dataset*, optou-se por utilizar a linguagem de programação Python® para manipulação do conjunto de dados e aplicação dos modelos de redes neurais, devido a facilidade de escrita da linguagem, possibilidade de se trabalhar com *notebooks* e ainda possuir diversas bibliotecas já implementadas e validadas que auxiliam o desenvolvimento de aplicações de AI.

A seguir, com a linguagem configurada, o *dataset* foi submetido a um processo de *split*, ou seja, foi dividido em dois conjuntos de dados, um para treinamento das redes neurais e outro para teste. Para isso, utilizou-se a biblioteca *scikit-learn*⁴ que disponibiliza um método voltado à separação de conjuntos de dados. Para este trabalho, a porcentagem de dados destinados ao treinamento foi de 80% (1136 imagens) e ao teste foi de 20% (285 imagens).

A partir disso, o caminho escolhido para realização dos testes e avaliação dos resultados baseou-se em similaridade vetorial, e para isso, o banco de dados vetorial *Qdrant*® se mostrou o mais robusto em alguns aspectos como os métodos disponibilizados, velocidade de consulta, facilidade de implementação e capacidade de armazenamento. Então, utilizando o *qdrant-client*⁵, biblioteca de conexão direta do Python® com o *Qdrant*®, criou-se uma *collection* no banco de dados, responsável por armazenar os vetores que representam o conjunto de treinamento.

Além disso, no momento de criação da *collection* é definido o algoritmo de busca que o banco de dados irá utilizar para realizar a consulta por similaridade baseado em um vetor de imagem. Para este trabalho, utilizou-se a métrica de similaridade do cosseno. Os vetores foram gerados a partir de modelos de rede neurais pré-treinados capazes de extrair de formas distintas as características de cada imagem e armazená-las de forma numérica em um vetor denso (*embedding*).

⁴ <https://scikit-learn.org/stable/>

⁵ <https://github.com/qdrant/qdrant-client>

Os modelos utilizados foram instalados de diferentes fontes para realização deste trabalho, incluindo modelos disponibilizados pela MetaAI® através no modelo pré-treinado DINO-V2 e também pela biblioteca em Python® *img2vec_pytorch*⁶. Além disso, para efeitos de testes iniciais também utilizou-se o Roboflow⁷, uma plataforma *web* voltada ao desenvolvimento de projetos de processamento de imagem e visão computacional. A

A biblioteca *img2vec_pytorch* oferece diversas arquiteturas de CNNs que podem ser utilizadas. As estruturas escolhidas para realização dos testes possuem diferentes dimensionalidades, ou seja, os vetores densos (*embeddings*) que são gerados possuem quantidades de posições diferentes para cada rede neural. Por exemplo, uma rede Resnet-18 utilizada gera um vetor de uma imagem com 512 posições, já uma rede Alexnet retorna um vetor de tamanho 4096. Em suma, todas as arquiteturas utilizadas para este trabalho estão dispostas na Tabela 2 a seguir, sendo as duas últimas referentes ao DINO-V2 e as demais ao *Img2Vec*.

Tabela 2 - Modelos de rede neural e suas respectivas dimensionalidades

Modelo	Dimensionalidade
Resnet-18	512
Resnet152	2048
Vgg-11	4096
Alexnet	4096
Densenet	1024
efficientnet_b0	1280
efficientnet_b7	2560
DINO-V2-Base	768
DINO-V2-Large	1521

Fonte: Elaborado pelo autor (2024)

Para cada um dos modelos utilizados, uma *collection* foi criada para armazenar os vetores gerados pelo respectivo modelo. Com todas as *collections* criadas a partir do conjunto de treinamento, o conjunto de testes foi utilizado para

⁶ https://github.com/christiansafka/img2vec/blob/master/img2vec_pytorch/img_to_vec.py

⁷ <https://roboflow.com/>

obter as acurácias nas classificações com base na comparação de semelhança entre os vetores armazenados e os vetores do conjunto de teste.

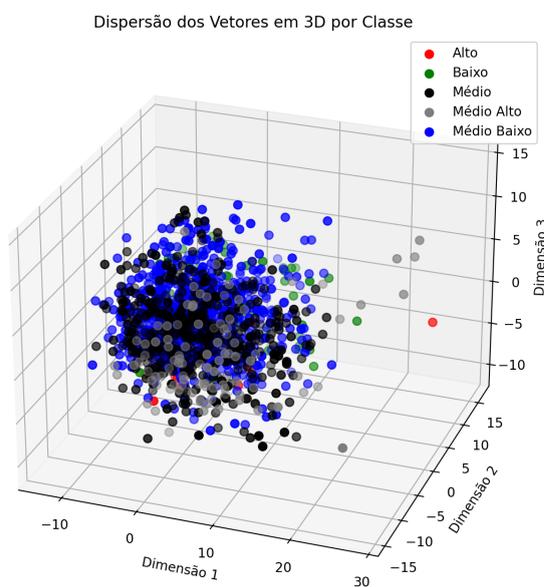
5.3 ANÁLISE E DISCUSSÃO DOS RESULTADOS

5.3.1 Detalhamento do *Dataset*

Para melhor compreensão dos resultados faz-se necessário um detalhamento inicial do *dataset* utilizado neste trabalho. Este é composto apenas por imagens de fachadas e suas respectivas classificações dos padrões construtivos, nesse sentido, a melhor maneira de visualizar como as imagens e suas classes estão dispostas é através de um gráfico de dispersão.

Dessa forma, a partir dos vetores gerados para cada imagem, reduziu-se a dimensionalidade para três dimensões de modo que fosse possível produzir a representação gráfica de onde cada imagem está espacialmente postada. A Figura 15 a seguir apresenta a disposição de cada imagem em três dimensões onde cada cor representa uma classificação diferente.

Figura 15 - Representação de todas as classes em 3D

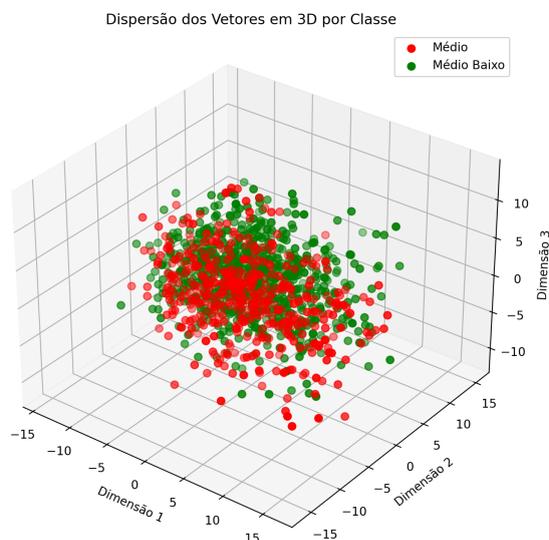


Fonte: Elaborado pelo autor (2024)

A partir da análise do gráfico é perceptível um desbalanceamento do *dataset*, onde aproximadamente 85% das imagens estão dispostas entre as classes Médio e

Médio Baixo, sendo 53.7% para a classe Médio Baixo e 32% para Médio. As Figuras 16 e 17 a seguir mostram mais visivelmente esse desbalanceamento.

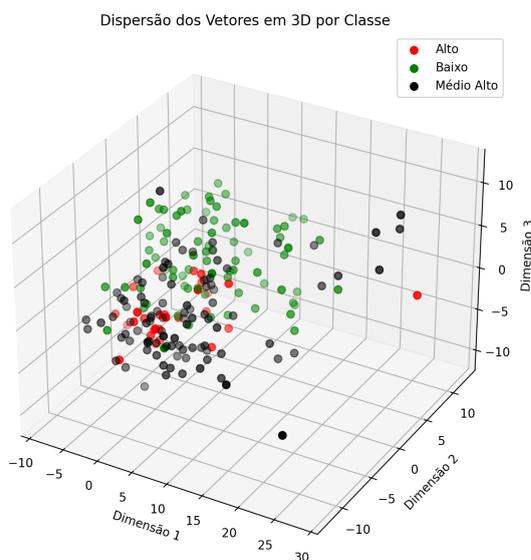
Figura 16 - Representação das classes Médio e Médio Baixo em 3D



Fonte: Elaborado pelo autor (2024)

Visualizando apenas as classes que representam 85% do total do *dataset* pode-se perceber que além da grande quantidade, os pontos estão muito próximos no espaço. O fato de estarem muito próximos no espaço pode influenciar as redes neurais utilizadas no momento de distinguir as classes para realizar uma classificação. Isto pode implicar na realização de classificações equivocadas, visto que o desbalanceamento de classes pode impactar no resultado final.

Figura 17 - Representação do restante das classes em 3D



Fonte: Elaborado pelo autor (2024)

A Figura 17 apresenta a disposição do restante das classes que representam 15% do *dataset*. Visualiza-se no gráfico também uma aproximação entre os pontos que representam a classe Médio Alto e Alto, o que também pode dificultar a compreensão das redes neurais. Além disso, por conta dessas três classes representarem uma parcela menor em relação ao total de imagens, a classificação pode ter uma tendência em prover resultados que apontem para as classes Médio e Médio Baixo. Considerando o retorno dessas classes, classificações imprecisas podem ser encontradas, visto que em um retorno do banco de dados com muitos vetores, há chances da maior parcela de vetores retornados serem de classe Médio ou Médio Baixo.

A partir do detalhamento dos dados e compreensão de como estão dispostos no espaço, objetiva-se promover maior clareza às análises dos resultados obtidos. Assim, a seção 5.3.2 a seguir, expõe como cada um dos modelos de rede neural testados desempenhou na classificação levando-se em conta o *dataset* de teste.

5.3.2 Análises Gerais dos Resultados

Após a análise do *dataset* tem-se elementos que permitem uma clareza maior sobre a avaliação dos resultados gerais. Assim, para se obter a acurácia para todos os modelos, um conjunto de passos foi realizado. Primeiramente, duas variáveis de base foram utilizadas, k e n , onde k representa a quantidade de classes sendo analisada e n a quantidade de vetores retornados do banco de dados a cada consulta. Dessa forma, tem-se k variando de 1 à 5 e n variando de 1 à 100 com os seguintes valores para o teste [1, 10, 25, 50, 75, 100].

Cada imagem do conjunto de testes foi submetida a uma função (Figura 18) que verifica a quantidade de registros de cada classe retornada com base na combinação dos valores de k classes e n imagens (vetores que constam no banco de dados vetorial). A partir do retorno, os vetores são agrupados por classe, e se a classe com a maior contagem corresponder à classificação correta da imagem avaliada (imagem de entrada), o resultado é considerado positivo, ou seja, um resultado correto; caso contrário, é considerado negativo, ou seja, um resultado incorreto. Para cada variação de k e n , os resultados positivos e negativos são contabilizados e armazenados.

Figura 18 - Função de histograma para cada variação de k e n

```

1 def histogram(classes:list, k:int):
2     result = {}
3     ctr=1
4     for value, key in sorted(((classes.count(e), e) for e in set(classes)), reverse=True):
5         if (ctr > k): break
6         result[key] = value
7         ctr+=1
8     return result

```

Fonte: Elaborado pelo autor (2024)

O retorno da função apresentada acima é a representação da quantidade de cada classe retornada dentre os n vetores recuperados. Por exemplo, para uma imagem de entrada com a classificação “Média”, utilizando $k=5$, que representa a quantidade total de classes e $n=100$, que representa o retorno de imagens do banco de dados com vetores mais próximos da imagem de entrada e ordenando pela quantidade decrescente, obtém-se o resultado da Tabela 3. Enfatizando, a coluna “Quantidade” representa o total de vetores retornados e está ligada à variável n .

Tabela 3 - Classificação para uma imagem de classe “Média” do conjunto de testes

Classe	Quantidade
4 - Médio Baixo	52
3 - Médio	46
5 - Baixo	1
2 - Médio Alto	1

Fonte: Elaborado pelo autor (2024)

No caso apresentado na Tabela 3 acima, como a classe de entrada é designada pelo rótulo “Médio” e o modelo retornou em maior quantidade vetores da classe “Médio Baixo”, isso representa um resultado negativo caso k seja 1, mas um resultado positivo caso, por exemplo, o k fosse igual a 2. Ou seja, se a classe da imagem de entrada estiver na primeira posição considerando $k=1$, o resultado seria

computado como negativo, do contrário, se estiver na segunda posição considerando $k=2$, o resultado seria computado como positivo. Após a contabilização de positivos e negativos, a acurácia é calculada para cada variação de k e n utilizando a Equação 1. Posteriormente, todas as acurácias foram calculadas e constam no Apêndice I.

Para a discussão geral dos resultados deste trabalho utilizou-se como base $k=2$ e $n=10$. Ainda que a escolha de k possa parecer arbitrária, esta se justifica devido às características do dataset e do cenário de estudo. Pode-se afirmar que o objetivo final da implementação do método é, para além da classificação do padrão construtivo de determinada imagem, também a recomendação de outras classes mais prováveis visando auxiliar a tomada de decisão do especialista.

Neste sentido, um $k=2$ provê alternativas para a escolha e definição do padrão construtivo correto que, ao longo do tempo, pode trazer maior assertividade às classificações. Ademais, k maiores que 2, como pode ser observado nos resultados do Apêndice I, promovem pequenas variações na acurácia. Conforme apresentado na Tabela 4, pode-se perceber que as acurácias totais ficaram entre 85.61% (pior caso) e 95.43% (melhor caso). Basicamente, este resultado é esperado e ocorre pelas características dos modelos pré-treinados baseados no DINO-V2, visto que foram treinados com uma quantidade de imagens quase 10 vezes maior em relação aos demais modelos testados. Isto promove um melhor mapeamento das imagens no espaço n dimensional, possibilitando classificações mais adequadas.

Tabela 4 - Acurácia de todos os modelos para $k=2$ e $n=10$

Modelo	Acurácia (%)
Alexnet	89,47%
Densenet	87,01%
efficientnet_b0	87,01%
efficientnet_b7	85,61%
resnet-18	91,12%
restnet152	89,47%
vgg	87,71%
DINO-V2-Base	95,43%
DINO-V2-Giant	95,08%

Fonte - Elaborado pelo autor (2024)

Para além dos modelos testados, também utilizou-se o *Roboflow*[®], uma ferramenta *web* para criação e treinamento de modelos de redes neurais de maneira visual. A ferramenta permite a integração com várias linguagens de programação e a criação de um *dataset* próprio, disponibilizando recursos como classificação, segmentação e reconhecimento de imagens. Seguindo a documentação da própria plataforma, o mesmo *dataset* de treinamento foi submetido às tecnologias do *Roboflow*[®] e um modelo de rede neural foi treinado utilizando esse conjunto de dados. Após o modelo treinado, este pode ser acessado via linguagens de programação para testes. O modelo obtido a partir do *Roboflow*[®] obteve uma acurácia geral de 63% para $k=1$ e $n=1$, uma acurácia maior quando comparada a maioria dos modelos, mas inferior ao DINO-V2. Dessa forma, optou-se por não utilizá-lo para demais análises visto que só há a possibilidade de retorno para $k=1$ e $n=1$. Ou seja, o modelo retorna apenas a classificação mais próxima, o que reduz as possibilidades de aplicação pensando em cenários de recomendação de uma lista de classes, visando subsidiar de maneira mais adequada a decisão de determinado especialista. Ademais, os resultados para esta configuração foram inferiores aos obtidos pelo modelo DINO-V2.

5.3.3 Análises Específicas dos Resultados

Esta seção objetiva apresentar três cenários específicos considerando os dados obtidos a partir do modelo de rede neural que apresentou os melhores resultados entre todos os modelos testados que constam na Tabela 2, considerando $k=1$ ou 2 e $n=10$.

Neste caso, a arquitetura que correspondeu de maneira mais eficaz aos testes de acurácia realizados foi o DINO-V2-Base que obteve uma acurácia de 95.43%. A partir dessa análise, evidencia-se que o modelo base do DINO-V2 possui uma dimensionalidade de 768 como representado na Tabela 2, a segunda menor entre todos os modelos analisados. Mesmo possuindo um vetor de menor tamanho, o DINO-V2-Base apresentou um desempenho superior em comparação com os outros modelos.

Esse resultado pode ser atribuído ao fato de que, quanto maior a dimensionalidade de um vetor, mais características são extraídas da imagem e,

quanto mais características são extraídas, mais complexa se torna a tarefa de classificação. No cenário atual, muitas imagens possuem baixa qualidade. Desta forma, ao se extrair mais características pode-se criar dificuldades para que um classificador indique corretamente determinada classe, visto que ao aprofundar na extração de características de uma região da imagem, estas podem se sobrepor de tal maneira que se parecem semelhantes quando não são. Nesse sentido, a dimensionalidade mais baixa do DINO-V2-Base contribuiu para uma representação mais eficaz das características das imagens, resultando em um melhor desempenho na tarefa de classificação. A Tabela 5 contém todos os resultados obtidos pelo melhor modelo, que também constam no Apêndice I.

Tabela 5 - Matriz de acurácias para cada variação de k e n do modelo DINO-V2-Base

k	$n = 1$	$n = 10$	$n = 25$	$n = 50$	$n = 75$	$n = 100$
1	62,10	68,77	69,12	68,77	69,47	68,42
2	62,10	95,43	95,08	94,73	94,73	94,73
3	62,10	97,89	98,59	99,29	99,29	98,94
4	62,10	97,89	98,59	100,00	100,00	100,00
5	62,10	97,89	98,59	100,00	100,00	100,00

Fonte: Elaborado pelo autor (2024)

Em sequência, visando uma representação mais visual dos resultados, serão discutidos três cenários distintos. Para o primeiro cenário (Figura 19) tem-se uma classificação positiva. Já para o segundo cenário (Figura 20) tem-se uma classificação negativa para $k=1$ e positiva para $k=2$. Por fim, o terceiro cenário (Figura 21) apresenta uma classificação incorreta para ambos $k=1$ e 2. Cada um desses casos foi analisado para um retorno de 100 imagens, ou seja, $n=100$. Posteriormente, para cada imagem que representa determinado cenário, é possível visualizar as respectivas tabelas de ranqueamento, que apresentam a porcentagem (%) de relevância de cada classe ordenada de maneira decrescente, onde a(s) classificação(ões) considerada(s) é/são determinada(s) pelo k .

Figura 19 - Foto de fachada com classe “Média” utilizada no primeiro cenário



Fonte: Obtido com a prefeitura de Criciúma

Tabela 6 - Ranking de classes retornado para o primeiro cenário

Posição	Classe	Quantidade	Relevância (%)
1º	3 - Médio	57	57%
2º	4 - Médio Baixo	35	35%
3º	2 - Médio Alto	4	4%
4º	5 - Baixo	3	3%
5º	1 - Alto	1	1%

Fonte: Elaborado pelo autor (2024)

A imagem acima (Figura 19) foi utilizada a partir do conjunto de testes e representava a classe “Médio” com identificador 3. Este cenário apresenta claramente um acerto, visto que para as 100 imagens mais próximas retornadas, 57% possuem a classificação “Médio” (Tabela 6), sendo a classificação considerada como um resultado positivo.

Já no cenário seguinte (Figura 20) representado a seguir o modelo apresenta um resultado negativo para $k=1$, visto que a maior relevância não representa devidamente a classe correta da imagem de entrada. Por outro lado, com $k=2$ o resultado seria considerado positivo.

Figura 20 - Foto de fachada com classe “Média” utilizada no segundo cenário



Fonte: Obtido com a prefeitura de Criciúma

Tabela 7 - Ranking de classes retornado para o segundo cenário

Posição	Classe	Quantidade	Relevância (%)
1º	4 - Médio Baixo	46	46%
2º	3 - Médio	38	38%
3º	2 - Médio Alto	10	10%
4º	5 - Baixo	3	3%
5º	1 - Alto	1	1%

Fonte: Elaborado pelo autor (2024)

Neste caso, a classe correta da imagem designada como “Média” (Tabela 7) e considerando $k=1$, foi disposta na segunda posição com 38% de relevância enquanto a classe com a maior relevância foi a “Médio Baixo” com 46%. Este resultado pode ser explicado em parte pelo desbalanceamento do *dataset*. Outro ponto a considerar é que o conjunto de dados, em muitos casos, apresentam classificações que parecem equivocadas, o que irá conduzir a erros de classificação. De qualquer modo, a proximidade das relevâncias para as duas primeiras posições sugere que o classificador teve dificuldades em discernir

adequadamente entre os padrões. E, desta forma, é fundamental a recomendação e escolha adequada pelo especialista.

O terceiro e último cenário apresentado (Figura 21) ilustra um erro de classificação, onde a classe correta da imagem aparece com uma relevância de apenas 5% (Tabela 8). A imagem de entrada foi rotulada como "Alta", no entanto, a maior relevância obtida para esta imagem foi para a classe "Médio", com 56%. Este caso exemplifica bem o desbalanceamento do dataset, onde a quantidade muito baixa de imagens classificadas como "Alta" impediu que o modelo obtivesse casos suficientes para superar a predominância da classe "Médio", que representa uma parcela significativamente maior no *dataset*.

Figura 21 - Foto de fachada com classe "Alta" utilizada no terceiro cenário



Fonte: Obtido com a prefeitura de Criciúma

Tabela 8 - Ranking de classes retornado para o terceiro cenário

Posição	Classe	Quantidade	Relevância (%)
1º	3 - Médio	56	56%
2º	2 - Médio Alto	21	21%
3º	4 - Médio Baixo	17	17%
4º	1 - Alto	5	5%
5º	5 - Baixo	1	1%

Fonte: Elaborado pelo autor (2024)

Os cenários apresentados demonstram o desafio do desbalanceamento de dados na classificação de imagens, especialmente quando se trabalha com classes de objetos que apresentam características semelhantes. As soluções para esse problema incluem técnicas de amostragem aleatória, ponderação de classes, técnicas de aumento de dados (do inglês *data augmentation* - DA) e métodos de sintetização de dados, visando garantir que os modelos de classificação sejam treinados com um conjunto de dados mais representativo e equilibrado, possibilitando resultados mais precisos e confiáveis.

6 CONSIDERAÇÕES FINAIS E TRABALHOS FUTUROS

O presente trabalho investigou a importância dos padrões construtivos com impacto, por exemplo, na arrecadação tributária de impostos sobre imóveis urbanos, destacando a relevância do emprego de geotecnologias na gestão imobiliária municipal, buscando se adequar às definições da ODS 11 que prevê até 2030 cidades mais sustentáveis. Mais especificamente, apresentou um método voltado à classificação e recomendação de padrões construtivos de residências, colaborando para a definição de políticas mais eficientes por governos municipais. Este estudo se baseou no uso da metodologia *Design Science Research Methodology* (DSRM) para o desenvolvimento de um método destinado a otimizar a classificação/recomendação de padrões construtivos. Ressalta-se ainda, o valor de aplicar uma abordagem multifacetada que combina pesquisa aplicada, inovação tecnológica e revisão bibliográfica para conceber soluções concretas, com potencial de aplicação prática.

O método proposto é constituído de seis etapas, desde a obtenção do *dataset* até a proposição de um classificador utilizando diversos modelos pré-treinados de redes neurais. De maneira geral, os resultados são promissores, sendo que o melhor resultado foi alcançado com o modelo *DinoV2-Base*, que obteve uma acurácia de 95,43% para a recomendação de duas classes ($k=2$). Para detalhar o funcionamento do método proposto e demonstrar os desafios enfrentados devido ao desbalanceamento presente no *dataset*, foram apresentados três cenários distintos na seção 5.3.3, demonstrando a aplicabilidade e potencialidade.

Além disso, o método proposto pode representar um possível mecanismo facilitador para os órgãos do governo responsáveis pelo cadastro imobiliário das cidades, auxiliando na promoção do desenvolvimento urbano. Com a materialização em tecnologia, é possível não apenas aprimorar a eficiência na classificação dos padrões construtivos, mas também possibilitar um processo de avaliação mais rápido e acurado. Isto beneficia a gestão tributária e contribui para uma administração pública mais transparente e eficaz para os cidadãos, como foco em cidades que sejam mais sustentáveis e forneçam um ambiente de qualidade para se viver, consonante ao ODS 11. Por fim, este trabalho demonstra potencial de aplicação real, combinando diferentes áreas que podem impactar na gestão urbana.

6.1 TRABALHOS FUTUROS

Adicionalmente, este trabalho busca fornecer uma base teórica e tecnológica pensando no desenvolvimento de novas técnicas e geotecnologias que vão de encontro aos temas propostos pelo ODS 11, de modo que os recursos gerados possam auxiliar em programas de governo e, assim, serem capazes de tornar as cidades mais inteligentes.

Para tal, uma possível continuidade para este projeto seria o desenvolvimento de um aplicativo móvel que utilize o modelo de rede neural para realizar a classificação e recomendação automatizada. Neste cenário, o funcionário responsável pelo cadastro do padrão construtivo poderia enviar uma foto da fachada através do aplicativo e receber recomendações sobre qual rótulo mais se adequa à imagem enviada. A decisão então seria armazenada em um banco de dados permitindo que o modelo de classificação tivesse um aprendizado contínuo.

Outro trabalho futuro possível reside na investigação de novas abordagens para as técnicas utilizadas, como por exemplo, o armazenamento de apenas um *embedding* representando a média dos vetores de cada classe de padrão construtivo. Dessa forma, o processamento realizado pelo classificador poderia ser agilizado, pois o vetor de entrada seria comparado com apenas cinco *embeddings* médios, e não com vários vetores armazenados no banco de dados. Isso poderia resultar em uma melhora significativa na velocidade de classificação e recomendação.

Além dessas implementações, a base de dados deste projeto poderia ser enriquecida com informações adicionais sobre as construções analisadas, como sua geolocalização e outras características físicas que interferem diretamente no valor estimado do imóvel. Sendo assim, o método não só definiria o padrão construtivo, mas também poderia fornecer uma estimativa do valor venal da construção. Isso incluiria dados como área construída, idade do edifício, material de construção, e proximidade de serviços essenciais. Existindo tal possibilidade, o trabalho de cálculo realizado pelas prefeituras seria significativamente otimizado, permitindo uma avaliação mais precisa e eficiente dos valores imobiliários, melhorando a arrecadação de impostos e a gestão urbana.

Concluindo, a continuidade dos estudos e o aprofundamento nos temas propostos neste trabalho podem oferecer contribuições substanciais tanto para a

academia quanto para a formulação de estratégias eficazes por parte dos gestores públicos no sentido de melhorar, não apenas a arrecadação tributária e a fiscalização, mas também o planejamento urbano e o desenvolvimento socioeconômico.

REFERÊNCIAS

ASSOCIAÇÃO BRASILEIRA DE NORMAS TÉCNICAS. NBR 6022. **NBR 12721:** Avaliação de custos de construção para incorporação imobiliária e outras disposições para condomínios edifícios. Rio de Janeiro, p. 61. 2005.

ABDURRAHMAN, L; MULYANA, T. Parallel Construction of Information Technology Value Model: Design-Science Research Methodology. **In: 8th International Conference on Information and Communication Technology (IColCT)**, Yogyakarta, Indonesia, 2020.

AHMED, I; JEON, G; PICCIALLI, F. From Artificial Intelligence to Explainable Artificial Intelligence in Industry 4.0: A Survey on What, How, and Where. **IEEE Transactions On Industrial Informatics**, VOL. 18, NO. 8, AUGUST 2022.

BECK, C. A. M. R; BOFF, M. M. Cidades Inteligentes: desigualdades, gentrificação e os desafios da implementação dos ODS. **Revista De Direito Econômico E Socioambiental**, Curitiba, v. 13, n.3, p.565-593, set./dez. 2022.

BOKHARI, S. A. A; MYEONG, S. The Influence of Artificial Intelligence on E-Governance and Cybersecurity in Smart Cities: A Stakeholder's Perspective. **IEEE Access**. v. 11, 10 July 2023.

BORTOLOTI, F.D; TAVARES, J; RAUBER, T.W; CIARELLI, P. M; BOTELHO, R. C. G. An annotated image database of building facades categorized into land uses for object detection using deep learning. **Machine Vision and Applications**, v. 33, n. 80, 2022.

BUSKO, M; APOLLO, M. Public Administration and Landowners Facing Real Estate Cadastre Modernization: A Win-Lose or Win-Win Situation? **Resources 2023**, v. 12, n. 6, p. 73, 2023.

CUGURULLO, F; CAPROTTI, F; COOK, M; KARVONEN, A; M^cGUIRK, P; MARVIN, S. The rise of AI urbanism in post-smart cities: A critical commentary on urban artificial intelligence. **Urban Studies**, v. 61(6), 2024.

DINI, S. F; WIBOWO, E. P; IQBAL, M; BAHAR, Y. N; ALFIANDY, A. Applying Deep Learning and Convolutional Neural Network System to Identity Historic Buildings: The "Little China" Building in Central Java, Indonesia. **ISVS e-journal**, v. 10, 2023.

ELNGAR, A. A; ARAFA, M; FATHY, A; MOUSTAFA, B; MAHMOUD, O; SHABAN, M; FAWZY, N. Image Classification Based On CNN: A Survey. **Journal of Cybersecurity and Information Management (JCIM)**, v. 6, n. 1, p. 18-50, 2021.

ENGSTRÖM, E; STOREY, M. A; RUNESON, P; HÖST, M; BALDASSARRE, M. T. How software engineering research aligns with design science: a review. **Empir Software Eng**, v. 25, 2630–2660, 2020.

FEDERICA, G; STEFFEN, M; ABDELGHANI, M; VOLKER, O. Building Stock Classification Using Machine Learning: A Case Study for Oslo, Norway. **Frontiers in Earth Science**, v. 10, 2022.

FREIRE, C. A; FERREIRA, F. A; CARAYANNIS, E. G; FERREIRA, J. J. Artificial Intelligence and Smart Cities: A DEMATEL Approach to Adaptation Challenges and Initiatives. **IEEE Transactions on Engineering Management**. v. 70, p. 1881 - 1899, 2021.

GERHARDT, T. E; SILVEIRA, D. T. Métodos de pesquisa. **Porto Alegre: Editora da UFRGS**, 2009.

GRAZZIOTIN, L. S; KLAUS, V; PEREIRA, A. P. M. Pesquisa documental histórica e pesquisa bibliográfica: focos de estudo e percursos metodológicos. **Pro-Posições**. v. 33, Campinas, SP, 2022.

HEVNER A; CHATTERJEE S. Design Research in Information Systems Theory and Practice. **Integrated Series in Information Systems**. v. 22, 2010.

HOSSAIN, A; SAJIB, S. A. Classification of Image using Convolutional Neural Network (CNN). **Global Journal of Computer Science and Technology**. v. 19, 2019.

INCITTI, F; URLI, F; SNIDARO, L. Beyond word embeddings: A survey. **Information Fusion**, v. 89, 2023.

JANIESCH, C; ZSCHECH, P; HEINRICH, K. Machine learning and deep learning. **Electron Markets** **31**, 685–695, 2021.

JUNIOR, V. F; WOSZEZENKI, C; ANDERLE, D. F; SPERONI, R; NAKAYAMA, M. K. A pesquisa científica e tecnológica. **Espacios**. v. 35, n. 9, p. 12, 2014.

KATTENBORN, T; LEITLOFF, J; SCHIEFER, F; HINZ, S. Review on Convolutional Neural Networks (CNN) in vegetation remote sensing. **ISPRS Journal of Photogrammetry and Remote Sensing**. v. 173, p. 24-49, 2021.

KIELA, D; BOTTOU, L. Learning Image Embeddings using Convolutional Neural Networks for Improved Multi-Modal Semantics. **Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)**, pages 36–45, 2014.

KOZLOWSKI, A. C; TADDY, M; EVANS, J. A. The Geometry of Culture: Analyzing the Meanings of Class through Word Embeddings. **American Sociological Review**, v. 84, 2019.

LAUPHEIMER, D; TUTZAUER, P; HAALA, N; SPICKER, M. Neural Networks For The Classification Of Building Use From Street-View Imagery, **ISPRS Ann. Photogramm. Remote Sens. Spatial Inf. Sci**, v. 2, 2018.

LECUN, Y; BENGIO, Y; HINTON, G. Deep learning. **Nature** **521**, 436–444, 2015.

LEI, X; PAN, H; HUANG, X. A Dilated CNN Model for Image Classification. **IEEE Access**, v. 7, pp. 124087-124095, 2019.

LIMA, T. C. S; MIOTO, R. C. T. Procedimentos metodológicos na construção do conhecimento científico: a pesquisa bibliográfica. **Rev. Katál. Florianópolis**. v. 10, p. 37-45, 2007.

MANDELBAUM, A; SHALEV, A. Word Embeddings and Their Use In Sentence Classification Tasks. **Hebrew University of Jerusalem**. 2016.

MASKELIŪNAS, R.; KATKEVIČIUS, A.; Plonis, D.; SLEDEVIČ, T.; MEŠKĖNAS, A.; DAMAŠEVIČIUS, R. Building Façade Style Classification from UAV Imagery Using a Pareto-Optimized Deep Learning Network. **Electronics**, v. 11, 2022

NATH, S. S; MISHRA, G; KAR, J; CHAKRABORTY, S; DEY, N. A survey of image classification methods and techniques. **International Conference on Control, Instrumentation, Communication and Computational Technologies (ICICCT)**, 2014.

NGUYEN, G. *et al.* Machine Learning and Deep Learning frameworks and libraries for large-scale data mining: a survey. **Artif Intell Rev** **52**, 77–124 (2019).

OSPINA, T. G; MORALES, Z. F. Gestión territorial y sus implicaciones con el ODS 11 Reflexiones desde Colombia y Costa Rica. **Revista de Arquitectura (Bogotá)**, v. 22, n. 2, p. 141–152, jul/dez. 2020.

OTTONI, A.L.C; DE AMORIM, R. M; NOVO, M. S; COSTA, D. B. Tuning of data augmentation hyperparameters in deep learning to building construction image classification with small datasets. **Int. J. Mach. Learn. & Cyber**, v. 14, 2023.

ÖZTÜRK, C; TAŞYÜREK, M; TÜRKDAMAR, M. U. Transfer learning and fine-tuned transfer learning methods' effectiveness analyse in the CNN-based deep learning models. **Concurrency Computat Pract Exper**. v. 35, 2023.

PAIVA, C. dos A.; ANTUNES, A. F. B. Geração De Planta De Valores Genéricos A Partir Do Cadastro Territorial Urbano. **Revista Brasileira de Cartografia**, v. 69, n. 3, 2017.

PEFFERS, K; TUUNANEN, T; ROTHENBERGER, M. A; CHATTERJEE, S. A Design Science Research Methodology For Information Systems Research. **Journal of management information systems : JMIS**, v. 24, n. 3, p. 45–77, 2007.

PENNINGTON, J; SOCHER, R; MANNING C. D. GloVe: Global Vectors for Word Representation. **Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)**. pages 1532–1543, 2014.

PIIRAINEN, K. A.; BRIGGS, R.O. Design Theory in Practice – Making Design Science Research More Transparent. **Service-Oriented Perspectives in Design Science Research**. v. 6629, 2011.

PIMENTEL, M; FILIPPO, D; SANTOS, T. M. Design Science Research: pesquisa científica atrelada ao design de artefatos. **Revista de Educação a Distância e Elearning**. v. 3, n. 1, março/abril 2020.

POUYANFAR S. et al. A Survey on Deep Learning: Algorithms, Techniques, and Applications. **ACM Computing Surveys**. v. 51, n. 5, Article 92, september 2018.

ROBERTSON, S; AZIZPOUR, H; SMITH, K; HARTMAN, J. Digital image analysis in breast pathology—from image processing techniques to artificial intelligence. **Translational Research**, v. 194, p. 19-35, 2018.

RODRIGUES, H. L. A; BORGES, M. G; LEITE, M. E. Geotecnologias Livres no Cadastro Imobiliário de Municípios do Norte de Minas Gerais. **Geo UERJ**, Rio de Janeiro, n. 39, 2021.

SABELLI, C. How the brain finds patterns in images. **Nature Italy**. 2022.

SAMHOURI, M; AI-ARABIAT, L; AL-ATRASH, F. Prediction and measurement of damage to architectural heritages facades using convolutional neural networks. **Neural Comput & Applic**, v. 34, 2022.

SARVAMANGALA, D. R; KULKARNI, R. V. Convolutional neural networks in medical image understanding: a survey. **Evol. Intel**. v. 15, p. 1–22, 2022.

SCHMIDHUBER J. Deep learning in neural networks: An overview. **Neural Networks**. v. 61, p. 85-117 2015.

SEJNOWSKI TJ. The unreasonable effectiveness of deep learning in artificial intelligence. **Proc Natl Acad Sci U S A**. v. 117, 2020 Dec 1.

SHARIFI, A; ALLAM, Z; BIBRI, S. E; KHAVARIAN-GARMSIR, A. R. Smart cities and sustainable development goals (SDGs): A systematic literature review of co-benefits and trade-offs. **Cities**, v. 146, 2024.

TAOUFIQ, S.; NAGY, B.; BENEDEK, C. HierarchyNet: Hierarchical CNN-Based Urban Building Classification. **Remote Sens**, 2020.

ULLAH, A; ANWAR, S. M; LI, J; NADEEM, L; MAHMOOD, T; REHMAN, A; SABA, T. Smart cities: the role of Internet of Things and machine learning in realizing a data-centric smart environment. **Complex & Intelligent Systems**, v. 10, 2024.

VIEIRA, J. A; LEITE, A. R; KUHN, A. S. Perspectivas da Produção de Pesquisa Aplicada, Inovação e Desenvolvimento Científico e Tecnológico nos Institutos Federais. **Revista Valore**. v. 8, 2023.

VICTORINO P. da S; AMORIM A; SHIMABUKURO M. H. Compartilhamento de Informações entre os Sistemas de Cadastro e de Registro de Imóveis utilizando um SIG Cadastral. **Revista Brasileira de Cartografia**. Rio de Janeiro, n. 67/6, p. 1119-1130, Set/Out/2015.

WANG, P; FAN, E; WANG, P. Comparative analysis of image classification algorithms based on traditional machine learning and deep learning. **Pattern Recognition Letters**. v. 141, p. 61–67, 2021.

WEVERS, M; SMITS T. The visual digital turn: Using neural networks to study historical images. **Digital Scholarship in the Humanities**, v. 35, n. 1, 2020.

APÊNDICE I

As tabelas a seguir apresentam, para os valores de k e n , resultados percentuais de acurácia.

Alexnet

k	$n = 1$	$n = 10$	$n = 25$	$n = 50$	$n = 75$	$n = 100$
1	55,78	62,10	62,45	57,54	58,59	59,64
2	55,78	89,47	88,42	88,07	87,71	87,01
3	55,78	94,03	97,54	97,19	96,84	97,19
4	55,78	95,78	98,94	99,64	100,00	99,64
5	55,78	95,78	98,94	99,64	100,00	100,00

densenet

k	$n = 1$	$n = 10$	$n = 25$	$n = 50$	$n = 75$	$n = 100$
1	54,38	60,35	61,05	62,10	60	60,70
2	54,38	87,01	87,71	89,12	89,12	88,07
3	54,38	95,08	94,73	96,49	97,89	97,89
4	54,38	96,14	97,89	98,94	100,00	100,00
5	54,38	96,14	98,24	99,29	100,00	100,00

efficientnet_b0

k	$n = 1$	$n = 10$	$n = 25$	$n = 50$	$n = 75$	$n = 100$
1	51,57	55,08	55,43	57,19	55,43	55,78
2	51,57	87,01	87,19	87,01	87,01	87,01
3	51,57	91,57	94,73	95,08	95,43	96,14
4	51,57	94,03	96,84	97,89	97,89	98,94
5	51,57	94,03	96,84	98,94	100,00	100,00

efficientnet_b7

<i>k</i>	<i>n = 1</i>	<i>n = 10</i>	<i>n = 25</i>	<i>n = 50</i>	<i>n = 75</i>	<i>n = 100</i>
1	54,38	57,54	57,19	57,19	56,14	56,84
2	54,38	85,61	87,01	85,95	87,01	87,01
3	54,38	92,28	93,68	93,33	93,33	94,38
4	54,38	93,33	97,54	98,59	98,59	98,94
5	54,38	93,68	97,54	99,29	99,29	100,00

resnet-18

<i>k</i>	<i>n = 1</i>	<i>n = 10</i>	<i>n = 25</i>	<i>n = 50</i>	<i>n = 75</i>	<i>n = 100</i>
1	52,28	62,80	63,85	61,75	60,35	60
2	52,28	91,12	88,77	88,07	88,07	88,42
3	52,28	96,14	97,19	96,49	96,84	97,19
4	52,28	96,14	98,59	100,00	100,00	99,64
5	52,28	96,14	98,59	100,00	100,00	100,00

restnet152

<i>k</i>	<i>n = 1</i>	<i>n = 10</i>	<i>n = 25</i>	<i>n = 50</i>	<i>n = 75</i>	<i>n = 100</i>
1	51,28	63,50	59,64	60	60,70	60,35
2	51,28	89,47	89,12	89,47	88,77	88,77
3	51,28	94,38	97,19	97,54	97,54	98,24
4	51,28	94,73	98,59	99,64	99,64	99,64
5	51,28	94,73	98,94	99,64	100,00	100,00

vgg

<i>k</i>	<i>n = 1</i>	<i>n = 10</i>	<i>n = 25</i>	<i>n = 50</i>	<i>n = 75</i>	<i>n = 100</i>
1	51,92	65,26	63,15	62,45	62,10	62,45
2	51,92	87,71	89,82	89,12	88,77	88,77
3	51,92	94,38	96,49	97,54	97,54	97,54
4	51,92	96,14	97,89	99,64	100,00	100,00
5	51,92	96,14	98,24	99,64	100,00	100,00

dinov2 base

<i>k</i>	<i>n = 1</i>	<i>n = 10</i>	<i>n = 25</i>	<i>n = 50</i>	<i>n = 75</i>	<i>n = 100</i>
1	62,10	68,77	69,12	68,77	69,47	68,42
2	62,10	95,43	95,08	94,73	94,73	94,73
3	62,10	97,89	98,59	99,29	99,29	98,94
4	62,10	97,89	98,59	100,00	100,00	100,00
5	62,10	97,89	98,59	100,00	100,00	100,00

dinov2 giant

<i>k</i>	<i>n = 1</i>	<i>n = 10</i>	<i>n = 25</i>	<i>n = 50</i>	<i>n = 75</i>	<i>n = 100</i>
1	67,36	68,77	69,47	69,47	68,07	67,01
2	67,36	95,08	94,73	95,08	94,38	94,38
3	67,36	98,24	98,94	99,64	99,28	99,29
4	67,36	98,24	99,29	100,00	100,00	100,00
5	67,36	98,24	98,29	100,00	100,00	100,00

ANEXO I

DECLARAÇÃO

Eu, Lucas Cardoso Benedet, portador do RG 6.792.587 e CPF 095.348.979-57, aluno regularmente matriculado no curso de Tecnologias da Informação e Comunicação da Universidade Federal de Santa Catarina sob a matrícula 21200547, venho por meio desta, solicitar o acesso de dados das unidades cadastrais e fotos de fachada para a realização do estágio de TCC onde objetiva usar inteligência artificial para automatização da classificação do padrão construtivo das edificações para contribuir com a pesquisa e o desenvolvimento dos trabalho de cadastro imobiliário e para fins didáticos na produção do meu Trabalho de Conclusão de Curso (TCC).

O objetivo da utilização desses dados é a realização de estudos e análises necessários para a conclusão do TCC, cuja temática será reconhecimento de imagem para classificação de padrões construtivos. Este trabalho é parte integrante dos requisitos para obtenção do grau de Bacharel, sob a orientação do professor Alexandre Leopoldo Gonçalves.

Declaramos ainda que o gestor da empresa eTopocart apoia e autoriza a utilização desses dados para fins acadêmicos e de pesquisa, reconhecendo a importância do estudo e sua relevância para a área de geoprocessamento.

 Documento assinado digitalmente
LUCAS CARDOSO BENEDET
Data: 26/08/2023 17:15:23-0300
CPF: ***.348.979-57
Verifique as assinaturas em <https://x.ufsc.br>

Lucas Cardoso Benedet
Aluno

 Documento assinado digitalmente
Alexandre Leopoldo Gonçalves
Data: 26/08/2023 17:22:15-0300
CPF: ***.740.349-57
Verifique as assinaturas em <https://x.ufsc.br>

Alexandre Leopoldo Gonçalves
Professor Orientador/UFSC

 Documento assinado digitalmente
JONATHAN JURANDIR CAMPOS
Data: 06/07/2024 12:25:38-0300
Verifique em <https://validar.it.gov.br>

Jonathan Campos
Gestor da Empresa eTopocart