

ESTUDO DE VIABILIDADE PARA CONSTRUÇÃO DE UM ALGORITMO DE DESCRIÇÃO DE IMAGENS NO CONTEXTO DE TURISMO INCLUSIVO

Alline Gomes Silva*

Alexandro Lima Damasceno**

Ruan Dos Santos Gondim***

RESUMO

Este estudo explora maneiras de promover a inclusão de turistas com deficiência visual no setor do turismo. O foco da proposta é a implementação de um algoritmo de descrição de imagens, com o intuito de proporcionar uma experiência imersiva. O trabalho compreende uma revisão abrangente da literatura que explora os desafios enfrentados por indivíduos com deficiência visual no contexto turístico, bem como as tecnologias atualmente disponíveis. A pesquisa está organizada em seções detalhadas, começando pela metodologia adotada, passando pela discussão dos resultados obtidos e culminando em uma conclusão que destaca a importância do desenvolvimento contínuo do turismo inclusivo. A inclusão de uma base teórica sobre os desafios enfrentados pelas pessoas com deficiências visuais, combinada com a exploração de conceitos de Inteligência Artificial, fortalece a relevância e a contribuição do presente estudo. Três estudos destacados na geração automática de legendas para imagens são apresentados, consolidando conhecimentos e identificando oportunidades específicas para a aplicação proposta. Testes pós-treinamento validam a eficiência do modelo, evidenciando a capacidade da Inteligência Artificial em compreender e descrever pontos turísticos. Destaca-se a abordagem codificador-decodificador pela sua pertinência ao turismo inclusivo, enfatizando a flexibilidade e adaptabilidade do modelo. Além de sua contribuição para a pesquisa em tecnologia assistiva, este trabalho estabelece uma base de soluções inovadoras, apresentando um princípio de soluções para promover ativamente a inclusividade no setor turístico. Representa, assim, um avanço significativo na construção de um turismo mais aberto e acessível, alinhado com as demandas de uma sociedade inclusiva e diversificada.

Palavras-chave: Turismo Inclusivo. Descrição de Imagens. Inteligência Artificial no Turismo.

* Autora, Graduada em Ciência da Computação, Instituto Federal de Educação, Ciência e Tecnologia do Ceará, Aracati, CE, Brasil. E-mail: allinegs80@gmail.com

** Orientador, Mestre em Ciência da Computação, docente do Instituto Federal de Educação, Ciência e Tecnologia do Ceará, Aracati, CE, Brasil. E-mail: alexandro.lima@ifce.edu.br

*** Coorientador, Mestre em Ciência da Computação, docente do Instituto Federal de Educação, Ciência e Tecnologia do Ceará, Aracati, CE, Brasil. E-mail: ruan.gondim@ifce.edu.br

ABSTRACT

This study explores ways to promote the inclusion of visually impaired tourists in the tourism sector. The focus of the proposal is the implementation of an image description algorithm with the aim of providing an immersive experience. The work encompasses a comprehensive literature review that explores the challenges faced by individuals with visual impairments in the tourism context, as well as the currently available technologies. The research is organized into detailed sections, starting with the adopted methodology, moving on to the discussion of the obtained results, and culminating in a conclusion that highlights the importance of continuous development in inclusive tourism. The inclusion of a theoretical foundation on the challenges faced by people with visual impairments, combined with the exploration of Artificial Intelligence concepts, strengthens the relevance and contribution of this study. Three highlighted studies on automatic image captioning are presented, consolidating knowledge and identifying specific opportunities for the proposed application. Post-training tests validate the model's efficiency, highlighting the capability of Artificial Intelligence to understand and describe tourist attractions. The encoder-decoder approach is emphasized for its relevance to inclusive tourism, emphasizing the flexibility and adaptability of the model. In addition to its contribution to assistive technology research, this work establishes a foundation for innovative solutions, presenting a principle of solutions to actively promote inclusivity in the tourism sector. It represents a significant advancement in building a more open and accessible tourism, aligned with the demands of an inclusive and diverse society.

Keywords: Inclusive Tourism. Image Descriptions. Artificial Intelligence in Tourism.

1 INTRODUÇÃO

O turismo é uma atividade universalmente apreciada, capaz de transcender barreiras culturais, geográficas e sociais (REIS, 2007). No entanto, muitos indivíduos com deficiências visuais são negligenciados dessa experiência sensorialmente rica que o turismo proporciona (ANACLETO, 2014). A falta de acessibilidade no setor turístico tem sido uma limitação significativa para essas pessoas, impedindo-as de explorar e apreciar o mundo de maneira plena (DUARTE; PEREIRA, 2017). Nesse contexto, surge a necessidade de desenvolver soluções inovadoras que promovam a inclusão de todos os viajantes, independentemente de suas capacidades visuais (VALVERDE; OLIVEIRA, 2022).

Conforme estipulado pelas diretrizes das Nações Unidas, é um requisito para os países membros assegurar que as pessoas com deficiência desfrutem de igualdade de oportunidades no que diz respeito ao acesso a atividades de lazer. Este princípio abrange uma ampla gama de experiências, como a capacidade de participar de restaurantes, cinemas, teatros, bibliotecas e outros locais de entretenimento, bem como a possibilidade de acessar estádios esportivos, hotéis, praias e outras instalações recreativas. A eliminação de barreiras para a plena inclusão é de responsabilidade dos países membros. Além disso, é essencial que as entidades relacionadas ao turismo, agências de viagens, organizações voluntárias e outros envolvidos na organização de atividades recreativas e oportunidades de viagem ofereçam serviços sem qualquer forma de discriminação contra as pessoas com deficiência. A acessibilidade e a inclusão devem ser princípios norteadores em todas as iniciativas relacionadas ao lazer e ao turismo (FERNANDES; PINHEIRO, 2013).

Diante dos pontos expostos, este trabalho propõe uma abordagem centrada na implementação de um algoritmo de descrição de imagens, concebido para promover a inclusão no turismo. Essa ferramenta de acessibilidade oferecerá informações detalhadas sobre imagens que variam desde obras de arte até paisagens naturais, permitindo uma maior imersão no turismo.

A relevância deste estudo concentra-se em uma questão crucial de inclusão que tem sido negligenciada por um longo período. À medida que a sociedade avança em direção a uma compreensão mais profunda da diversidade e da igualdade de oportunidades, torna-se imperativo que o setor do turismo se ajuste para atender às necessidades de todos os seus potenciais clientes. Nesse contexto, o estudo de viabilidade não apenas diminuirá as barreiras enfrentadas por pessoas com deficiência visual, mas também fomentará um ambiente mais inclusivo e acolhedor para todos os viajantes (FRAGATA et al., 2019).

Em um mundo cada vez mais conectado e consciente da importância da inclusão, este trabalho busca abrir novas perspectivas no campo do turismo, onde a visão vai além das limitações e se torna verdadeiramente inclusiva. Através do desenvolvimento e implementação do algoritmo, espera-se criar um ambiente em que todos possam compartilhar igualmente as riquezas do turismo, independentemente de suas restrições visuais (KITA, 2022).

Na seção 2, apresenta-se uma revisão da literatura, discutindo os desafios enfrentados pelas pessoas com deficiência visual no turismo e as tecnologias existentes para melhorar a

acessibilidade. Na seção 3, descreverá as metodologias, resultados obtidos e relevâncias dos trabalhos de outros autores relacionados com o tema. Na seção 4, será descrito o modelo e arquitetura utilizada, juntamente com o *dataset* e o treinamento descrito. Na seção 5, serão apresentados os resultados obtidos, incluindo testes e limitações encontradas. Por fim, na seção 6, serão discutidas as conclusões do trabalho, suas implicações e as direções futuras para o desenvolvimento do turismo inclusivo.

2 FUNDAMENTAÇÃO TEÓRICA

Nesta seção, será abordada a fundamentação teórica que sustenta a compreensão dos obstáculos enfrentados por indivíduos com deficiência visual no contexto do turismo inclusivo. É apresentada a relevância da acessibilidade e da sensibilização como elementos-chave para fomentar uma experiência turística inclusiva(ANACLETO, 2014). Além disso, serão discutidos os princípios da Inteligência Artificial aplicados no desenvolvimento do algoritmo proposto para a descrição de imagens, visando facilitar a implementação do turismo acessível.

2.1 Acessibilidade no Turismo

A acessibilidade é um dos pilares fundamentais para garantir que pessoas com deficiência visual possam desfrutar de atividades turísticas(BIELER, 2020). Segundo a Convenção sobre os Direitos das Pessoas com Deficiência da ONU (2006), a acessibilidade é um direito fundamental, e os Estados signatários têm a responsabilidade de garantir que pessoas com deficiência tenham igualdade de oportunidades para participar na vida cultural, social e turística. Isso envolve, entre outras coisas, a adaptação de infraestruturas e serviços turísticos para atender às necessidades específicas desse público(ANDRADE, 2022).

2.2 Comunicação e Informação Acessível

A comunicação e a informação acessível são elementos críticos para garantir que pessoas com deficiência visual possam desfrutar de experiências turísticas de maneira completa(RODRIGUES; PERINOTTO, 2022). Isso envolve o fornecimento de informações de maneira que seja compreensível para todas as pessoas, independentemente de suas deficiências. Alguns dos métodos de comunicação e informação acessíveis incluem:

- (i) Braille: O Braille é um sistema de escrita tátil utilizado por pessoas cegas ou com baixa visão. Sinalizações, mapas e informações em Braille são essenciais em locais turísticos para permitir a orientação e a compreensão(LOCH, 2008).
- (ii) Áudio descrições: Para museus, exposições e locais turísticos, áudio descrições são narrações que descrevem visualmente elementos e detalhes, tornando a experiência mais rica para pessoas com deficiência visual(MATTOSO et al., 2012).

- (iii) *Websites*: *Websites* e aplicativos de turismo devem ser desenvolvidos com acessibilidade em mente, seguindo padrões como o *Web Content Accessibility Guidelines (WCAG)*. Isso inclui a disponibilização de alternativas textuais para imagens, navegação por teclado e outras funcionalidades que facilitam o acesso.(STRINGARI, 2021)

2.3 Sensibilização e Capacitação

A sensibilização e a capacitação são aspectos-chave para garantir que o setor de turismo atenda às necessidades das pessoas com deficiência visual. Os desafios que essas pessoas enfrentam no turismo são frequentemente agravados pela falta de conhecimento e sensibilização por parte dos prestadores de serviços turísticos (NASCIMENTO, 2018). A sensibilização e a capacitação envolvem:

- (i) *Treinamento de funcionários*: Funcionários em hotéis, restaurantes, atrações turísticas e outros locais devem receber treinamento sobre como atender de maneira sensível às necessidades de clientes com deficiência visual. Isso inclui orientações sobre como comunicar-se, como ajudar na orientação e como lidar com equipamentos de assistência (NASCIMENTO, 2018).
- (ii) *Promoção da inclusão*: A promoção da inclusão nas práticas de marketing e publicidade é importante para demonstrar o compromisso de um destino turístico ou empresa em receber visitantes com deficiência visual (CARDOSO, 2018).

2.4 Legislação e Regulamentações

A legislação e as regulamentações desempenham um papel fundamental no contexto do turismo acessível, garantindo que pessoas com deficiência tenham igualdade de oportunidades para participar de experiências turísticas(MOURA, 2013).

Convenção sobre os Direitos das Pessoas com Deficiência (CDPD): A CDPD, adotada pela Assembleia Geral da ONU em 2006, é um marco importante para a promoção dos direitos das pessoas com deficiência, incluindo o direito de participar na vida cultural, social e turística. A convenção estabelece a importância da acessibilidade em todos os aspectos da sociedade, incluindo o turismo(LOPES et al., 2009).

2.4.1 Legislação Nacional

De acordo com o último censo (2010) do Instituto Brasileiro de Geografia e Estatística (IBGE), o Brasil abriga mais de 6,5 milhões de pessoas com deficiência visual severa, sendo 582 mil cegas e 6 milhões com baixa visão. Apesar desse expressivo número, persistem casos de discriminação e preconceito. Decorrente a isso, a legislação nacional inclui:

- (i) Lei do Cão-Guia (Lei nº 11.126/2005): Essa lei garante o direito da pessoa com deficiência visual, acompanhada de cão-guia, de ingressar e permanecer com o animal em todos os meios de transporte e estabelecimentos abertos ao público.
- (ii) Lei de Cotas (Lei 8.213/1991): Estabelece a obrigatoriedade de empresas com 100 ou mais empregados destinarem de 2% a 5% de seus cargos a beneficiários reabilitados ou pessoas portadoras de deficiência.
- (iii) Lei de Braille (Lei nº 4.169/1962): Oficializa e obriga o uso do sistema Braille em todo o território nacional para escrita e leitura dos cegos.
- (iv) Lei de Isenção de IPI (Lei nº 8.989/1995): Garante isenção do Imposto sobre Produtos Industrializados para pessoas com deficiência física, visual, auditiva, mental severa, profunda ou transtorno do espectro autista na compra de automóveis específicos.
- (v) Lei de Acessibilidade (Decreto-lei nº 5.296/2004): Estipula prazos e regulamenta atendimento às necessidades específicas de pessoas com deficiência em projetos arquitetônicos, urbanísticos, de comunicação, informação e transporte coletivo.
- (vi) Plano Viver sem Limite (Decreto nº 7.612/2011): Institui o Plano Nacional dos Direitos da Pessoa com Deficiência, promovendo o exercício pleno e equitativo dos direitos, com foco em acesso à educação, atenção à saúde, inclusão social e acessibilidade.

2.5 Tecnologias Assistivas

A acessibilidade digital é uma preocupação crescente na sociedade contemporânea, especialmente quando se trata de garantir que todos tenham igualdade de acesso à informação e oportunidades. As pessoas com deficiências visuais, em particular, enfrentam desafios significativos ao tentar navegar na internet e utilizar dispositivos eletrônicos. Felizmente, várias tecnologias foram desenvolvidas para melhorar a acessibilidade, tornando possível que eles desfrutem de uma experiência digital mais inclusiva (SILVA, 2015). As tecnologias assistivas envolvem:

- (i) Leitores de Tela e Tecnologias de Síntese de Voz: Um dos avanços mais significativos na melhoria da acessibilidade para pessoas com deficiências visuais é a criação de leitores de tela e tecnologias de síntese de voz. Leitores de tela são softwares que interpretam e vocalizam o conteúdo exibido na tela, permitindo que os usuários com deficiência visual ouçam o que está sendo exibido. Além disso, as tecnologias de síntese de voz têm evoluído, tornando as vozes mais naturais e personalizáveis (MELO; COSTA; SOARES, 2006). Exemplos notáveis incluem: *JAWS (Job Access With Speech)* e *NVDA (NonVisual Desktop Access)*, bem como assistentes de voz como *VoiceOver* da *Apple* e *TalkBack* da *Google* (SILVA et al., 2021).
- (ii) Braille Digital e Linhas Táteis: Com o avanço da tecnologia, foram desenvolvidos dispositivos e aplicativos que permitem a leitura em Braille por meio de displays táteis. Além

disso, as linhas Braille digitais conectam-se a computadores e dispositivos móveis para traduzir o conteúdo exibido na tela em Braille em tempo real. Essas tecnologias facilitam a aprendizagem e a comunicação para pessoas com deficiências visuais, oferecendo-lhes uma maneira eficaz de interagir com dispositivos eletrônicos(FACCIONI, 2012).

- (iii) Reconhecimento de Voz e Comandos de Voz: O reconhecimento de voz avançado e os comandos de voz têm se tornado cada vez mais comuns. Essas tecnologias permitem que pessoas com deficiência visual controlem dispositivos e acessem informações por meio de comandos de voz. Assistentes virtuais como a Siri da *Apple*, a Alexa da *Amazon* e o *Google Assistant* tornaram-se parte integrante da vida cotidiana de muitas pessoas com deficiência visual(LIMA et al., 2015).
- (iv) Tecnologias de Navegação *Indoor* e *Outdoor*: A mobilidade é uma questão crítica para pessoas com deficiência visual. Tecnologias de navegação, como *GPS* adaptado e aplicativos específicos, auxiliam na orientação em ambientes internos e externos. Além disso, o desenvolvimento de tecnologias de mapeamento e sensores tem aprimorado a capacidade de detectar obstáculos e direcionar as pessoas com deficiências visuais de forma segura(WOHFART; ROQUE; SANTOS, 2015).
- (v) Aplicativos e Softwares de Acessibilidade: Muitos aplicativos e softwares foram desenvolvidos para melhorar a acessibilidade, desde aplicativos de leitura de livros falados até aplicativos de reconhecimento de objetos e cores. Essas ferramentas oferecem funcionalidades específicas para atender às necessidades de pessoas com deficiências visuais(SILVA; FREIRE, 2018).

2.5.1 *Inteligência Artificial (IA) e Turismo Inclusivo*

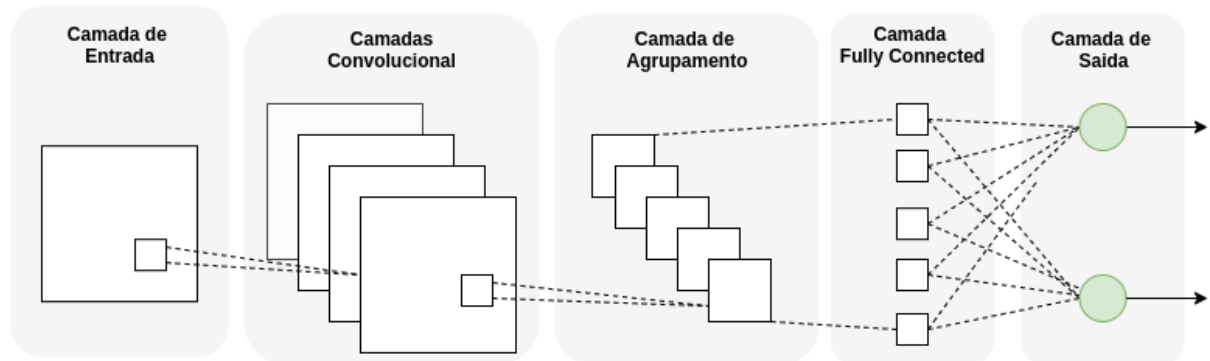
A Inteligência Artificial, campo interdisciplinar que combina ciência da computação, estatística e aprendizado de máquina, tem desempenhado um papel transformador em diversos setores. Por meio do desenvolvimento de algoritmos capazes de aprender padrões complexos a partir de dados, a IA possibilita a automação de tarefas, a análise preditiva e a resolução de problemas complexos. Sua capacidade de imitar a inteligência humana, tanto em termos de raciocínio quanto de processamento de dados, tem implicações significativas na otimização de processos e na criação de soluções inovadoras(JARRAHI, 2018).

No contexto do turismo inclusivo, a Inteligência Artificial assume um papel significativo como uma ferramenta capacitadora para promover a acessibilidade, especialmente no que se refere à experiência visual. O desenvolvimento de algoritmos para a descrição de imagens, impulsionados por IA, representa uma abordagem inovadora para enriquecer e tornar acessíveis as experiências turísticas, especialmente para pessoas com deficiência visual. Ao integrar algoritmos de processamento de linguagem natural e visão computacional à *softwares*, essa convergência têm o potencial de oferecer descrições detalhadas e precisas do ambiente ao redor, capacitando os usuários a explorar destinos turísticos de forma independente(TSAIH; HSU, 2018).

2.5.2 Redes Neurais Convolucionais (CNN)

Uma Rede Neural Convolutiva (*CNN*) é uma classe de redes neurais profundas que se mostrou eficaz na análise de imagens. Composta por uma camada de entrada (recebe a imagem de entrada), uma camada de saída (produz a saída final da rede) e várias camadas ocultas, como evidencia a Figura 1:

Figura 1 – Estrutura simplificada de uma rede neural convolutiva



Segurança em Redes 5G: Oportunidades e Desafios em Detecção de Anomalias e Predição de Tráfego Baseadas em Aprendizado de Máquina - Scientific Figure on ResearchGate.[accessed 16 Nov, 2023].

As camadas ocultas de uma *CNN* normalmente incluem: camadas convolucionais (responsáveis por realizar operações de convolução na imagem de entrada, utilizando filtros para extrair características locais); camadas de *pooling* (responsáveis por reduzir a dimensionalidade dos mapas de características, preservando características importantes); camadas totalmente conectadas (conectam todos os neurônios da camada anterior a cada neurônio da camada atual) (O'SHEA; NASH, 2015).

As *CNNs* têm diversas aplicações, incluindo reconhecimento de imagem e vídeo, sistemas de recomendação e processamento de linguagem natural. Têm se destacado em tarefas de processamento de imagens devido à sua capacidade de aprender características hierárquicas e espaciais (ALZUBAIDI et al., 2021).

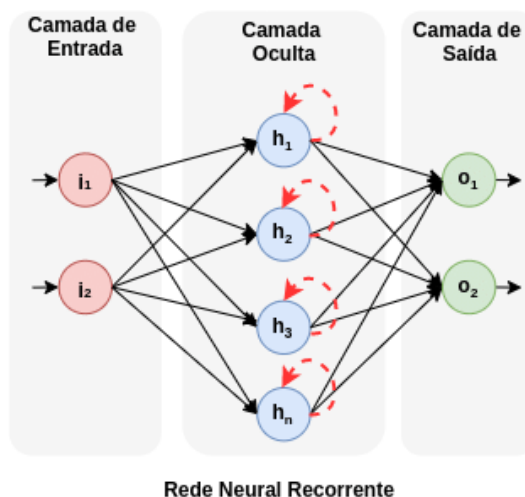
A capacidade da *CNN* para identificar padrões complexos em diferentes escalas a torna particularmente eficaz na extração de informações relevantes de imagens, sendo ideal para a tarefa de codificação em um contexto de descrição de imagens para turismo inclusivo (NOORD; POSTMA, 2017).

2.5.3 Redes Neurais Recorrentes (RNNs)

As Redes Neurais Recorrentes (*RNNs*) são um tipo especial de arquitetura de rede neural projetada para lidar com dados sequenciais ou temporais. Ao contrário das redes neurais tradicionais, as *RNNs* possuem uma estrutura que permite a retroalimentação de informações, permitindo que informações anteriores sejam consideradas na análise de dados atuais, como evidencia a Figura 2. Onde a camada de entrada recebe os dados sequenciais, a camada oculta

processa sequencialmente os dados (recebendo a entrada atual juntamente com a saída da etapa anterior) e por fim, a camada de saída que produz a saída final da rede (GROSSBERG, 2013).

Figura 2 – Interação entre os neurônios de uma Rede Neural Recorrente



Segurança em Redes 5G: Oportunidades e Desafios em Detecção de Anomalias e Predição de Tráfego Baseadas em Aprendizado de Máquina - Scientific Figure on ResearchGate.[accessed 16 Nov, 2023].

A capacidade de modelar dependências temporais faz com que as *RNNs* sejam particularmente eficazes em tarefas onde a ordem dos dados é crucial, como em séries temporais, linguagem natural e descrição de imagens em sequência (GRAVES et al., 2006).

2.6 Memória de Curto Prazo (*LSTM*)

A memória de curto prazo em redes neurais, refere-se a uma arquitetura de células projetada para lidar com o problema de gradientes que se desvanecem ou explodem em redes neurais recorrentes (*RNNs*) (STAUEMEYER; MORRIS, 2019). Os gradientes, referem-se às inclinações ou derivadas parciais das funções de perda em relação aos parâmetros do modelo. São utilizados no processo de treinamento para ajustar os pesos da rede, minimizando a função de perda. Ao propagar os gradientes ao longo de muitos passos temporais, eles podem diminuir drasticamente (desvanecer) ou aumentar exponencialmente (explodir), tornando o treinamento instável ou ineficaz (ARPIT et al., 2018).

A memória de curto prazo possui unidades de memória que são capazes de armazenar informações por longos períodos, tomando decisões sobre quais informações reter ou esquecer. Isso é alcançado através de portões de entrada, saída e esquecimento, que são componentes essenciais para controlar o fluxo de informações na célula da *LSTM* (STAUEMEYER; MORRIS, 2019).

Na descrição de imagens, são frequentemente empregadas para gerar sequências de palavras ou frases que descrevem o conteúdo visual. Isso é alcançado ao alimentar as representações da imagem através da Memória de Curto Prazo, que aprende a relação temporal e a dependência entre as palavras, criando descrições contextualmente relevantes e fluentes. A habilidade de lidar com dependências temporais em dados sequenciais torna-as cruciais para

tarefas de sequenciamento, incluindo a descrição de imagens, onde a ordem e a relação entre as palavras são fundamentais para comunicar efetivamente o conteúdo visual (WEERAKODY et al., 2021).

2.7 Mecanismo de atenção

O mecanismo de atenção no contexto de processamento de informações visuais e reconhecimento de padrões refere-se a uma abordagem computacional inspirada na maneira como os seres humanos direcionam sua atenção para partes específicas de uma cena visual. Este conceito é fundamental em campos como visão computacional e aprendizado de máquina, onde a capacidade de identificar elementos relevantes em uma imagem é crucial (TSOTSOS, 2021).

Assim como os seres humanos não processam todas as informações visuais de uma vez, os mecanismos de atenção em sistemas computacionais buscam concentrar seus recursos em partes específicas de uma imagem. Essa abordagem ajuda a reduzir a carga computacional, melhorar a eficiência e concentrar esforços nas regiões mais importantes da cena (TSOTSOS, 2021).

Em redes neurais, especialmente em modelos de aprendizado profundo, os mecanismos de atenção são incorporados para permitir que a rede dê mais peso a certas partes da entrada, tornando o processo de decisão mais adaptável e flexível. Os mecanismos de atenção podem ser adaptados para tarefas específicas (HERNÁNDEZ; AMIGÓ, 2021). No contexto do trabalho atual, pode ser utilizado para enfatizar elementos relevantes para uma pessoa com deficiência visual, como sinais indicativos, pontos de referência importantes, etc.

2.8 Avaliação - BLEU (*Bilingual Evaluation Understudy*)

O avanço da tecnologia de processamento de linguagem natural tem impulsionado a necessidade de avaliação rigorosa de sistemas de tradução automática. Essa avaliação tornou-se uma etapa crucial para garantir a eficácia e a precisão das aplicações que dependem da geração automática de texto, especialmente em contextos nos quais a comunicação é fundamental (WU et al., 2016).

No cenário específico do estudo proposto, a qualidade da tradução automática desempenha um papel central. Portanto, a precisão e a fluidez das descrições geradas automaticamente são essenciais para assegurar que as informações transmitidas sejam compreensíveis e enriquecedoras para todos os públicos (KINGHORN; ZHANG; SHAO, 2019). A tradução automática não se limita apenas à transposição de idiomas; ela também envolve a transmissão adequada de conceitos, emoções e nuances, particularmente quando se trata de descrever imagens que compõem o cenário turístico (DURET et al., 2023). Assim, a avaliação cuidadosa desses sistemas se torna fundamental para garantir uma descrição precisa.

Ao considerar o contexto do turismo inclusivo, a qualidade das descrições automáticas não apenas afeta a acessibilidade, mas também contribui diretamente para a experiência do usuário, influenciando a capacidade de formar uma conexão significativa com o ambiente explorado.

Portanto, a integração de métricas de avaliação, como o *BLEU*, torna-se imperativa para medir e aprimorar continuamente a eficácia do algoritmo.

3 TRABALHOS RELACIONADOS

Nesta seção, serão apresentados e discutidos três estudos significativos no campo da geração automática de legendas para imagens. Ao examinar esses trabalhos, busca-se consolidar conhecimentos que contribuem para a compreensão abrangente do estado atual do campo, identificando lacunas e oportunidades para o estudo de viabilidade na construção do algoritmo.

O estudo realizado (SURESH; JARAPALA; SUDEEP, 2022) se concentra na avaliação de modelos automáticos de geração de legendas de imagens (*NIC*) que empregam codificadores CNN e decodificadores *RNN*, como *LSTM* e unidades recorrentes com portas. A pesquisa foi conduzida no conjunto de dados *Flickr8k*, utilizando métricas qualitativas e quantitativas, como *ROUGE* e *BLEU-n*. Os resultados destacam que a combinação de um codificador *ResNet-101* e um decodificador *LSTM* supera em cenários específicos, como a concatenação condicional de *par-inject* e busca em feixe (*Beam Search*). Este estudo não apenas oferece *insights* sobre a otimização de modelos para tarefas específicas de processamento de imagens e linguagem natural, mas também serve como uma base valiosa para pesquisas futuras nesse campo, proporcionando uma compreensão aprofundada das diferentes arquiteturas e estratégias de decodificação.

Os autores (QU et al., 2019), apresentam uma abordagem inovadora para a geração de descrições de imagens, introduzindo a Rede de Memória de Curto Prazo com *Feedback* Baseado em Atenção (*AFLN*). Esta pesquisa visa aprimorar a semântica de alto nível e a precisão em algoritmos existentes, incorporando uma rede LSTM com *feedback* baseado em atenção durante a decodificação e utilizando o Bloco Convolutivo de Módulo de Atenção (*CBAM*) na fase de codificação. Os resultados nos dados COCO validam a eficácia da rede proposta, evidenciando melhorias substanciais em métricas como *BLEU-4* e *CIDEr*. Esta contribuição significativa oferece uma solução eficaz para melhorar a precisão e semântica em comparação com métodos existentes, estabelecendo uma base sólida para futuras pesquisas na área de processamento de imagens e geração de descrições.

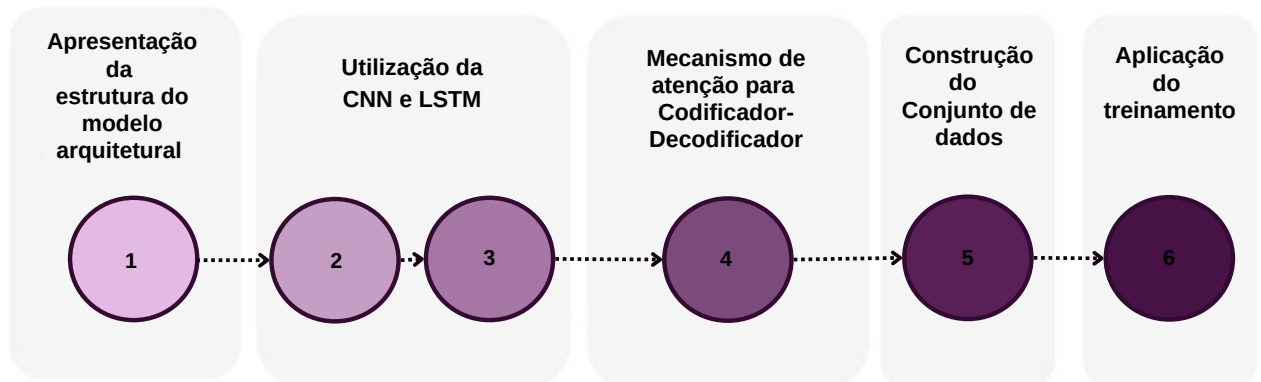
No contexto da geração de legendas de imagens, este artigo (TANTI; GATT; CAMILLERI, 2018) explora a disposição da informação da imagem em um modelo de linguagem de rede neural recorrente (*RNN*). Investigando duas abordagens - incorporação direta das características da imagem na *RNN* e fusão dessas características em uma camada pós-*RNN* - o estudo conclui empiricamente que a escolha entre essas arquiteturas não impacta significativamente no desempenho. No entanto, destaca que a arquitetura de fusão oferece vantagens práticas, reduzindo o tamanho do vetor de estado oculto da *RNN*. A pesquisa contribui empiricamente para a comparação sistemática entre essas abordagens, indicando que ambas são viáveis, mas a abordagem de fusão pode ser mais eficiente em termos de memória, fornecendo uma direção valiosa para futuros desenvolvimentos na geração de legendas de imagens.

4 METODOLOGIA

Esta seção irá apresentar as atividades de elaboração do algoritmo de descrição de imagens para turismo inclusivo, destacando cada componente do processo de desenvolvimento, incluindo o escopo do modelo, o conjunto de dados utilizado e a descrição do treinamento do algoritmo. Ao definir as estratégias empregadas, é ressaltado a robustez da arquitetura escolhida para o sistema alcançar resultados confiáveis e significativos, conforme delineado na seção 5.

A Figura 3 ilustra as seis etapas da metodologia utilizadas no desenvolvimento do projeto.

Figura 3 – Etapas da abordagem metodológica

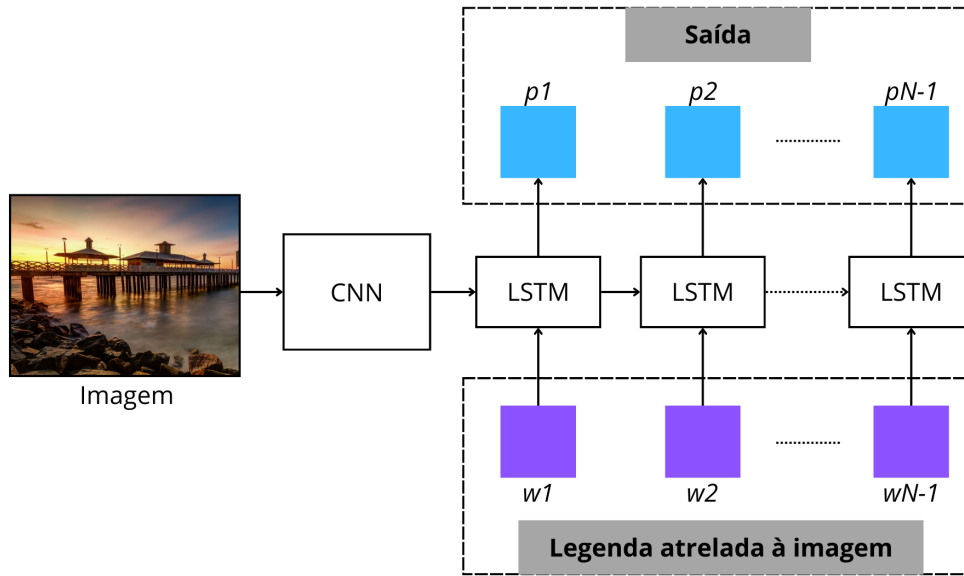


Fonte: Autor (2023)

No âmbito deste trabalho, foi escolhida uma arquitetura de codificador-decodificador, combinada juntamente com um mecanismo de atenção, para a construção do algoritmo de descrição de imagens. O codificador desempenha a função de extrair representações das imagens, enquanto o decodificador é encarregado de gerar legendas correspondentes.

O modelo adotado é fundamentado em uma rede neural, transitando diretamente de uma Rede Neural Convolutacional (*CNN*) para o processamento visual até uma Memória de Curto Prazo (*LSTM*) para a geração de linguagem. Este modelo realiza eficazmente a tarefa de gerar frases completas em linguagem natural a partir de uma imagem de entrada, conforme ilustrado na Figura 4, onde w seria uma palavra, $wN - 1$ a última palavra que consiste a legenda completa da imagem e p a palavra de saída do algoritmo, seguidamente por $pN - 1$ a última palavra que compõe a descrição final da imagem.

Figura 4 – Estrutura do Modelo



Autor (2023)

4.1 Rede Neural Convolucional (CNN) - Codificador

A aplicação da Rede Neural Convolucional (CNN) tem como objetivo a extração de um conjunto de vetores de características denominados “vetores de anotação”. Esses vetores, representados por L elementos são compostos por a_i , onde cada a_i é um vetor de características de dimensão D :

$$a = \{a_1, \dots, a_L\}, a_i \in \mathbb{R}^D \quad (1)$$

A extração de características ocorre em uma camada convolucional inferior, o que proporciona informações sobre partes específicas da imagem em duas dimensões. A abordagem possibilita que o decodificador concentre sua atenção de maneira seletiva em determinadas regiões da imagem, permitindo a escolha de um subconjunto específico de vetores de características para a geração da legenda. Essa seletividade direciona o foco do modelo para elementos relevantes da imagem durante o processo de decodificação da legenda.

4.2 Gerador de frases baseado em Memória de Curto Prazo (LSTM) - Decodificador

Na geração das legendas, como citada anteriormente, foi utilizada a Memória de Curto Prazo, onde cada palavra é produzida em passos temporais sucessivos. A abordagem adotada condiciona a geração de cada palavra com base em um vetor de contexto, no estado oculto anterior da rede e nas palavras já geradas. O vetor de contexto, que representa informações visuais da imagem, desempenha um papel crucial no processo de geração da legenda. Esse vetor é calculado dinamicamente para cada palavra, utilizando vetores de anotação associados a diferentes partes da imagem.

Para calcular o vetor de contexto, é incorporado um mecanismo de atenção, introduzindo um modelo que gera pesos indicando a importância relativa de diferentes partes da imagem na geração da próxima palavra. Esse método permite que o modelo leve em consideração informações contextuais específicas ao gerar cada palavra da legenda, adaptando-se à relevância de diferentes regiões da imagem durante o processo de geração textual. A capacidade de atenção contribui consideravelmente para a qualidade e contextualização da legenda gerada pelo modelo.

4.3 Mecanismo de Atenção

O mecanismo de atenção em cada passo do decodificador estabelece uma conexão direta com o codificador, permitindo que o decodificador concentre sua atenção em uma parte específica da imagem de origem. Essa abordagem possibilita ao decodificador direcionar sua atenção durante a geração de cada palavra, contribuindo para uma descrição mais precisa.

A dinâmica desse mecanismo é calculada individualmente para cada palavra da legenda a partir de vetores de anotação associados a diferentes partes da imagem. Através de um módulo chamado "*fat*", os pesos de atenção são calculados para cada vetor de anotação. Esses pesos indicam a importância relativa de diversas regiões da imagem no processo de geração da próxima palavra na legenda. Esse processo adaptativo e dinâmico de atenção é fundamental para melhorar a qualidade e a contextualização das palavras geradas durante a descrição da imagem.

O modelo opera ao receber uma única imagem em sua forma bruta e tem a capacidade de gerar uma legenda y codificada como uma sequência de palavras. Essas palavras são representadas por códigos 1-de- K , onde K representa o tamanho do vocabulário disponível para o modelo, C indica o comprimento total da legenda gerada e y_i representa um vetor associado à palavra i na sequência de palavras que compõem a legenda.

$$y = \{y_1, \dots, y_C\}, y_i \in \mathbb{R}^K \quad (2)$$

Em essência, o modelo traduz a informação visual da imagem em uma sequência estruturada de palavras, cada uma representada de maneira codificada em relação ao vocabulário definido.

Para a aplicação do mecanismo de atenção, é essencial ter um detalhamento do conjunto de dados utilizado. Pois influencia diretamente na capacidade do modelo de compreender e interpretar os elementos visuais da imagem. Dessa forma, contribui para o desempenho do modelo na tarefa de geração de descrições.

4.4 Conjunto de Dados

Baseou-se em informações disponibilizadas por um dos integrantes do projeto, cujo intuito era compilar um conjunto de dados diversificado. O conjunto de dados em questão é composto por 180 imagens de pontos turísticos localizados em diferentes cidades, sendo elas: Lisboa(PT), Coimbra(PT), Recife-PE(BR), Aracati-CE(BR), Fortaleza-CE(BR), Icapuí-CE(BR) e Itaiçaba-CE(BR).

Este conjunto de dados abrange uma ampla variedade de pontos turísticos, capturando a riqueza cultural e geográfica de diferentes regiões. A escolha dessas localidades estratégicas visa enriquecer a diversidade do *dataset*, oferecendo uma perspectiva abrangente de diferentes atrativos turísticos.

As imagens incluídas no conjunto de dados foram obtidas de fontes diversas, sendo cuidadosamente selecionadas para representar fielmente a beleza e a singularidade de cada local. A diversidade de pontos turísticos, que incluem desde monumentos históricos até paisagens naturais, proporciona um conjunto abrangente e enriquecedor para análises subsequentes. A seguir são apresentados fragmentos do *dataset* das cidades mencionadas anteriormente (4.4) nas Figuras 5,6,7,8,9,10,11:

Figura 5 – Monumento a Antero de Quental - Lisboa-PT



Fonte: Google Imagens

Figura 6 – Sé Velha de Coimbra - Coimbra-PT



Fonte: Google Imagens

Figura 7 – Colégio Damas - Recife(PE)



Fonte: Google Imagens

Figura 8 – Igreja Nossa Senhora dos Prazeres - Aracati(CE)



Fonte: Google Imagens

Figura 9 – Ponte dos Ingleses - Fortaleza(CE)



Fonte: Google Imagens

Figura 10 – Mirante do cruzeiro - Icapuí(CE)



Fonte: Google Imagens

Figura 11 – Igreja Matriz de Itaiçaba - Itaiçaba(CE)



Fonte: Google Imagens

Além das imagens, o *dataset* é complementado por 5 legendas associadas a cada fotografia. Numeradas de 0 à 4, essas legendas desempenham o papel de compreensão e contextualização das imagens, fornecendo informações relevantes sobre cada ponto turístico capturado. As legendas foram criadas manualmente pela autora do estudo e voluntários do IFCE - Campus Aracati, garantindo uma correlação precisa entre a descrição textual e o conteúdo visual. Um exemplo de legendas para uma imagem é apresentado na Figura 12:

Figura 12 – Exemplo de descrição da imagem



- (i) Um beagle e um golden retriever lutando na grama .
- (ii) Dois cães estão lutando na grama .
- (iii) Dois cachorrinhos estão brincando na grama verde .
- (iv) Dois cachorrinhos brincando na grama .
- (v) Dois cachorrinhos brincam na grama .

Fonte: Kaggle - Flickr8k Dataset (2023)

Em resumo, a metodologia empregada na criação e análise do *dataset* reflete um esforço meticuloso para garantir a diversidade e relevância dos dados. Essa abordagem proporciona um fundamento sólido para investigações futuras, promovendo uma compreensão aprofundada e significativa dos atrativos turísticos nas cidades selecionadas.

Além de servir como uma base sólida para análises subsequentes, o *dataset* desempenha um papel vital no treinamento do modelo. As imagens e legendas cuidadosamente escolhidas são os *inputs* essenciais para o processo de aprendizado da máquina, permitindo ao modelo capturar nuances e correlações entre a descrição textual e o conteúdo visual.

4.5 Treinamento

Após a construção do *dataset*, o próximo passo foi o treinamento do modelo proposto. Devido às restrições de memória e processamento pela ausência da assinatura do *Google Colab PRO*, a decisão foi tomada para adotar o *Kaggle* como a principal plataforma de compilação. Essa opção acarretou em um aumento no tempo necessário para configurar o ambiente de trabalho, uma vez que foram identificados desafios relacionados à importações de bibliotecas mais antigas. Além da escolha por não utilizar um acelerador de hardware no ambiente de execução, acarretando em um maior tempo de processamento.

Além do *dropout*, outra estratégia de regularização utilizada foi a parada antecipada no escore *BLEU*. A ideia por trás da parada antecipada é economizar tempo de treinamento, interrompendo o processo assim que não há mais melhoria observada em uma métrica específica

(nesse caso, o escore *BLEU*) ao longo de um número pré-definido de épocas. Como o *BLEU* é a métrica mais comumente relatada nas pesquisas relacionadas, foi escolhida para ser utilizada no conjunto de validação para seleção do modelo. Também valem ressaltar os parâmetros utilizados na Tabela 1:

Tabela 1 – Parâmetros de treinamento relevantes

Treinamento	
LR (Learning Rate)	3e-5
Número de épocas	15
Gradient clipping	5.
Teacher Forcing Ratio	0.
Batch Size	32
Função de perda	CrossEntropy
Otimizador	RMSProp

Fonte: Autor (2023)

A eficácia do treinamento do modelo depende significativamente dos parâmetros escolhidos. Abaixo estão os detalhes dos parâmetros relevantes utilizados durante o treinamento:

- (i) *Learning Rate (LR)*: 3e-5. O *Learning Rate* controla a magnitude dos ajustes nos pesos do modelo durante o treinamento. Um valor de 3e-5 indica uma taxa de aprendizado relativamente baixa, o que é comum em tarefas de processamento de linguagem natural.
- (ii) *Número de Épocas*: 15. O modelo foi treinado através de 15 épocas, representando 15 passagens completas pelo conjunto de treinamento. Esse número pode variar dependendo da complexidade da tarefa e da quantidade de dados disponíveis.
- (iii) *Gradient Clipping*: 5. A técnica de *Gradient Clipping* foi utilizada para evitar explosões no gradiente durante o treinamento. O valor de 5 indica que os gradientes foram cortados se excedessem esse limite.
- (iv) *Teacher Forcing Ratio*: 0. O *Teacher Forcing Ratio* define a proporção de exemplos em que as previsões corretas são alimentadas de volta ao modelo durante o treinamento. Com um valor de 0, o modelo não recebe previsões corretas como entrada nos passos subsequentes.
- (v) *Batch Size*: 32. O *Batch Size* determina o número de exemplos de treinamento usados em cada iteração do treinamento. Um valor de 32 é relativamente pequeno, mas pode ser necessário devido a restrições de memória.
- (vi) *Função de Perda*: *CrossEntropy*. A *CrossEntropy* é uma escolha comum para tarefas de classificação e foi utilizada para medir a discrepância entre as previsões do modelo e os rótulos reais.
- (vii) *Otimizador*: *RMSProp*. O *RMSProp* é um otimizador que ajusta adaptativamente a taxa de aprendizado para cada parâmetro. Sua escolha pode impactar a convergência e eficácia do treinamento.

Para uma visão mais ampla do treinamento, código-fonte disponível em: <<https://github.com/allinegsPY/Gerador-de-legenda>>

5 RESULTADOS

Esta seção destaca os principais resultados obtidos com a avaliação das métricas através da saída da rede do algoritmo envolvendo a análise de diversas categorias, proporcionando uma visão abrangente da capacidade da IA treinada para compreender e descrever a diversidade de pontos turísticos presentes no conjunto de dados.

Além disso, é importante ressaltar dois resultados significativos alcançados durante a avaliação. O primeiro destaca-se pela construção de uma base totalmente nova e inteiramente em português, demonstrando a capacidade do algoritmo em gerar conteúdo de forma autêntica e localizada. Essa característica contribui para a relevância e acessibilidade do sistema em contextos específicos. O segundo resultado notável é a descrição eficaz do algoritmo, evidenciando sua habilidade em descrever de maneira precisa e eficiente os pontos turísticos analisados. Essa capacidade de descrição aprimorada fortalece a utilidade prática da IA, oferecendo uma ferramenta valiosa para a organização, obtendo uma compreensão mais rápida e precisa da vasta variedade de locais turísticos presentes nos dados.

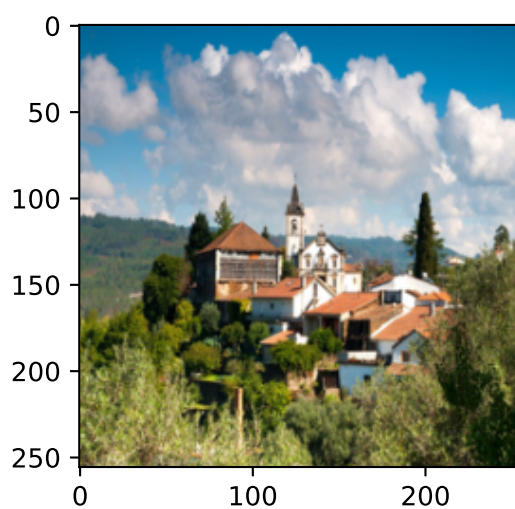
5.1 Resultados da Descrição de Imagens do Dataset

São apresentados três resultados obtidos nas descrições das imagens dos pontos turísticos selecionados. As Figuras (13, 15 e 14) destacam a eficácia da IA, onde “Resultado do algoritmo” seria a legenda gerada pelo treinamento e “Legendas verdadeiras” as legendas descritivas criadas manualmente para a imagem. Proporcionando *insights* adicionais sobre a capacidade do modelo em contextos práticos.

Figura 13 – Resultado do treinamento - Aldeia de Xisto (PT)

Resultado do algoritmo:
Vilarejo em colina cercado por árvores .

Legendas verdadeiras:
Diversas casas se sobressaem entre árvores .
Há casas de diferentes tamanhos em foco e ao fundo uma igreja com uma torre alta .
Casas de alturas e formas diferentes coexistem em uma zona rural .
Árvores rodeiam uma cidade pequena .
Casas diversas tem muros cobertos com plantas enquanto uma torre de igreja se destaca ao fundo .

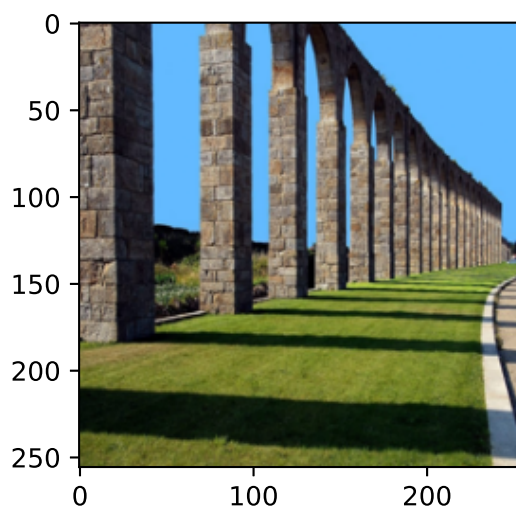


Fonte: Google Imagens

Figura 14 – Resultado do treinamento - Aqueduto de Santa Clara (PT)

Resultado do algoritmo:
Monumento alto de pedras arqueadas com grama verde .

Legendas verdadeiras
Aqueduto antigo com pilares, grama e um céu azul .
Construção de pedras arqueadas em grama verde .
Aqueduto antigo com arcos, campo gramado.
Pilares de pedras arqueados em grama verde .
Antigo aqueduto com arcos em área gramada .

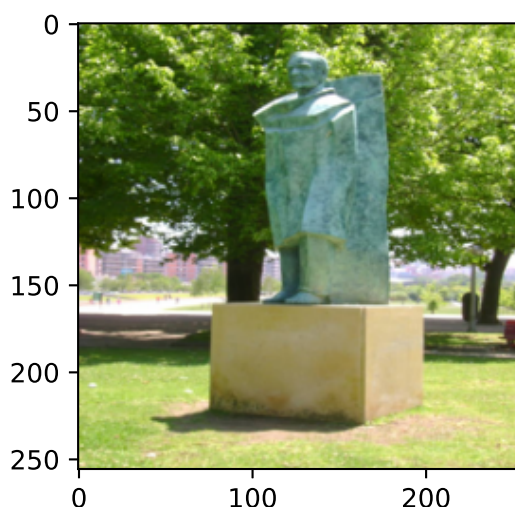


Fonte: Google Imagens

Figura 15 – Resultado do treinamento - Monumento Manuel Alegre (PT)

Resultado do Algoritmo:
Monumento verde sobre um bloco próximo de árvores .

Legendas verdadeiras:
Grande estátua verde de um homem no parque, com árvores ao fundo .
Estátua verde de um homem, colocado sobre um bloco .
Monumento verde em formato de um homem sobre um bloco amarelo em um parque .
Estátua no parque com árvores e bancos .
Estátua verde de um homem descalço em um pedestal .



Fonte: Google Imagens

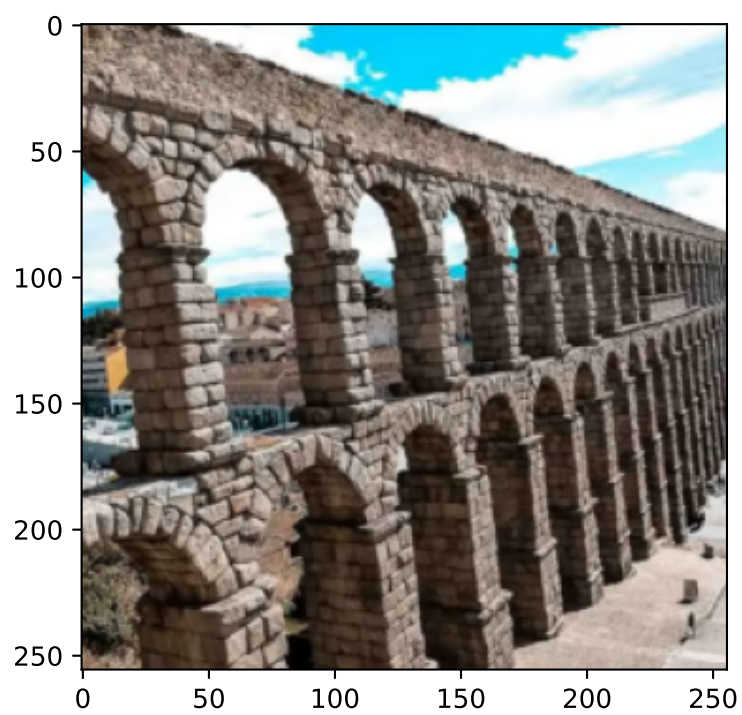
Como é possível verificar, a análise revela que as legendas produzidas alcançam notáveis proximidades com as legendas elaboradas por seres humanos, tanto em termos de compreensão, quanto de qualidade descritiva.

5.2 Resultados de Descrição de Amostras para Teste de Generalização do Algoritmo

Em prosseguimento, realizaram-se mais três testes do algoritmo, nos quais foram empregadas imagens previamente excluídas do conjunto de dados original: As Figuras (16, 17 e 18). De maneira notável, observou-se uma redução significativa na taxa de precisão durante essas avaliações adicionais. Este fenômeno levanta questionamentos pertinentes sobre a capacidade do algoritmo em generalizar para imagens fora do conjunto inicial de treinamento, destacando a necessidade de ajustes e otimizações para aprimorar sua robustez em diferentes contextos e cenários.

Figura 16 – Resultado do teste generalizado - Aqueduto de Segóvia (Espanha)

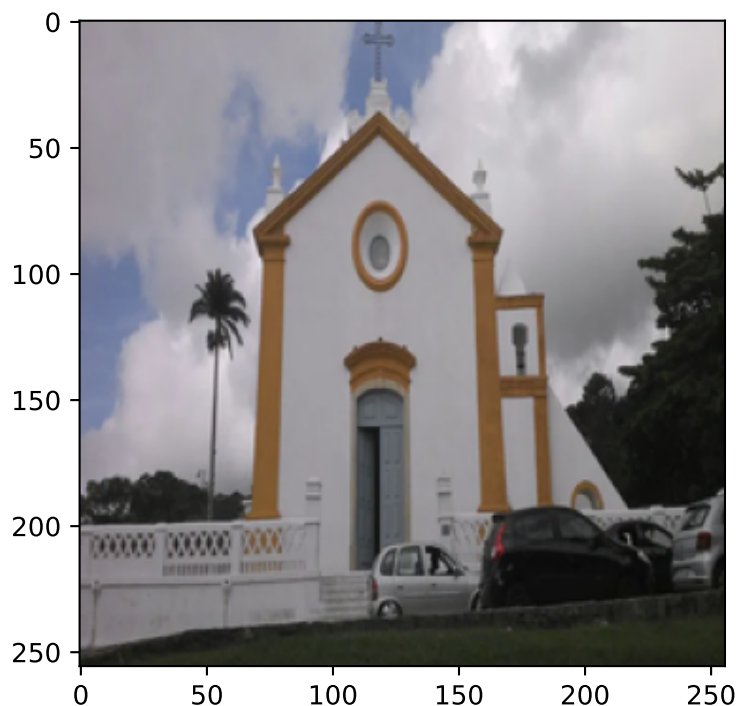
Resultado do algoritmo:
Muro de pedras circulares .



Fonte: Adriana/Em Algum Lugar do Mundo (2023)

Figura 17 – Resultado do teste generalizado - Igreja de Santo Antônio de Lisboa (Florianópolis-SC)

Resultado do algoritmo:
Igreja branca e <unk> <unk> .



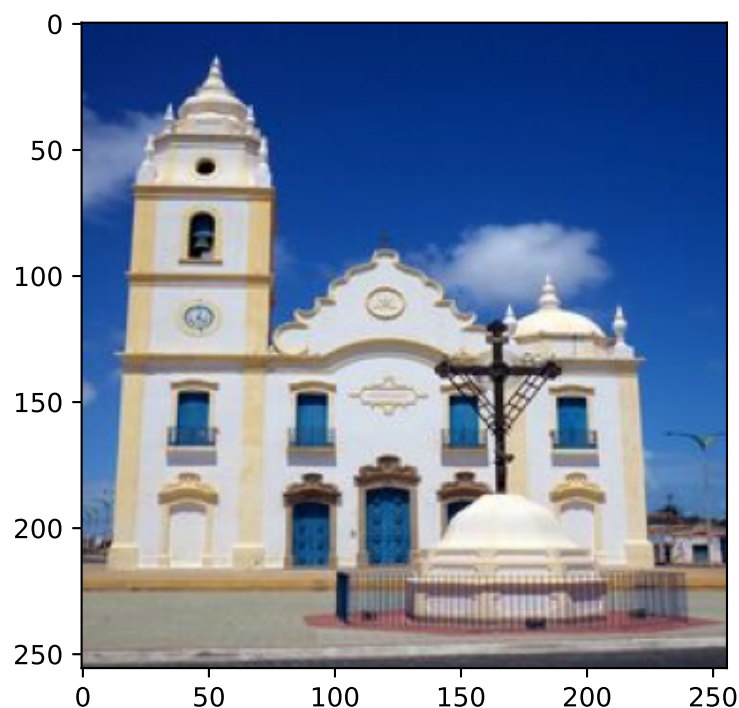
Fonte: Joana Caldas/G1 (2021)

Neste caso, o modelo não foi capaz de identificar as palavras que originalmente estavam presentes na sequência e substituiu essas palavras desconhecidas pelo marcador “<unk>”. A abreviação “<unk>” vem de “*unknown*” (desconhecido em inglês) e é empregada quando o modelo encontra uma palavra que não está presente em seu vocabulário pré-definido.

Quando um modelo de linguagem encontra uma palavra que não reconhece, em vez de tentar adivinhar ou usar um marcador específico para cada palavra desconhecida, ele simplesmente substitui a palavra desconhecida pelo marcador “<unk>”. O mesmo ocorreu no teste generalizado da Figura 18:

Figura 18 – Resultado do teste generalizado - Igreja Matriz de Nossa Senhora do Rosário (Aracati-CE)

Resultado do algoritmo:
Igreja branca com <unk> <unk> .



Fonte: GLandovsky (2014)

5.3 Desempenho da IA na Descrição de Imagens Turísticas

A avaliação do desempenho da Inteligência Artificial na descrição do conjunto de dados revelou uma taxa média de precisão de 85%. A análise de diferentes categorias, como arquitetura, paisagens naturais e cores, evidenciou a eficácia particular na entrega de descrições detalhadas. Como especifica a Tabela 2:

Tabela 2 – Descrição detalhada da detecção das categorias

Categoria de Descrição	Taxa de Precisão
Arquitetura	0.82691
Paisagens Naturais	0.88911
Cores	0.92078
Outros	0.90078

Fonte: Autor (2023)

5.3.1 Desenvolvimento e Treinamento do modelo

Para contextualizar o desempenho do modelo, é essencial analisar as métricas de treinamento. Quatro parâmetros foram monitorados: *Epoch*, *Loss*, *Accuracy* e *BLEU-4*.

- (i) *Epoch* (Época): É uma única passagem completa através de todo o conjunto de dados de treinamento durante o treinamento do modelo. O número de *Epochs* influencia diretamente a capacidade do modelo de aprender com os dados de treinamento.
- (ii) *Loss*: É uma medida que quantifica o quão longe as previsões do modelo estão dos rótulos reais. Uma queda consistente na *Loss* ao longo das épocas indica que o modelo está aprendendo efetivamente os padrões nos dados de treinamento.
- (iii) *Accuracy* (Precisão): É uma métrica fundamental que mede a proporção de previsões corretas em relação ao total de exemplos.
- (iv) *BLEU-4*: O *BLEU-4* considera até quatro palavras consecutivas, avaliando a sobreposição entre as sequências geradas pelo modelo e as sequências de referência. Um aumento consistente no *BLEU-4* ao longo das épocas indica que o modelo está melhorando na geração de saídas mais próximas às referências desejadas.

A Figura 19 apresenta o Treino e Validação no gráfico de involução do *Loss* por *Epochs* e seguidamente, a Figura 20 o gráfico da *Accuracy* e Avaliação *BLEU-4* também por *Epochs* durante o treinamento do modelo. Observa-se uma convergência eficaz, indicando que o modelo atingiu estabilidade durante o treinamento, resultando nos altos níveis de precisão observados na seção 5.3.

Figura 19 – Gráfico de Loss

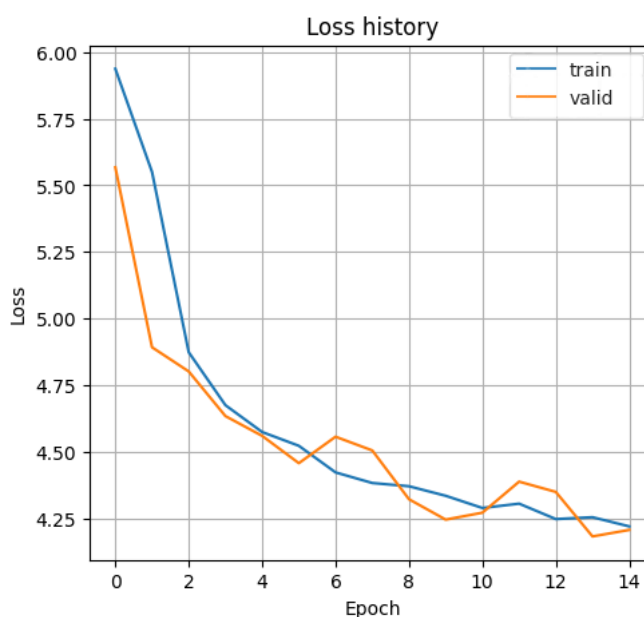
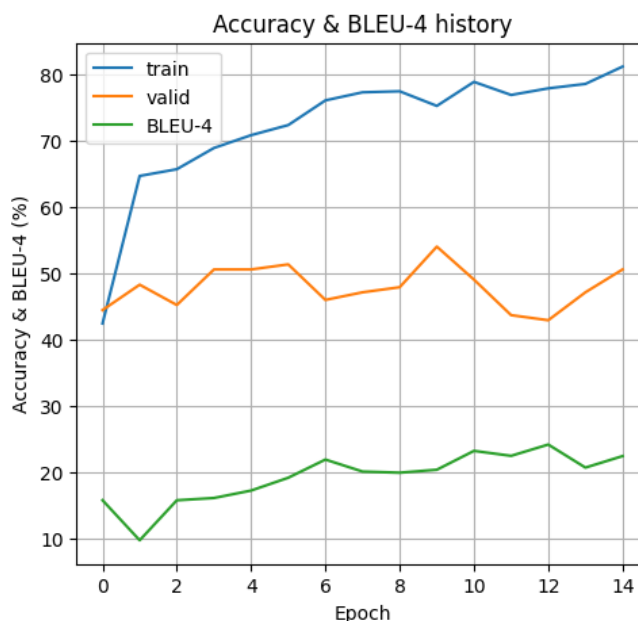


Figura 20 – Gráfico de *Accuracy* e *BLEU*

Autor (2023)

5.4 Limitações

Apesar do desempenho sólido da Inteligência Artificial, foram identificados desafios específicos em cenários caracterizados por iluminação inadequada, fotografias datadas e na aplicação da técnica de *Data Augmentation* no conjunto de dados. Notavelmente, durante a condução de testes adicionais com imagens não incluídas no conjunto original, observou-se uma redução na taxa de precisão, destacando a sensibilidade do modelo a variações nas condições de entrada.

A limitação da disponibilidade de um conjunto de dados mais abrangente emerge como um fator crítico a ser considerado. Este ponto ressalta a importância de um esforço contínuo para a expansão e diversificação do conjunto de dados de treinamento, a fim de fortalecer a capacidade da Inteligência Artificial em lidar com uma gama mais ampla de situações e desafios.

Este cenário enfatiza a necessidade constante de buscar melhorias, visando garantir a consistência nas descrições geradas, independentemente das condições ambientais. O reconhecimento dessas limitações não apenas evidencia a importância de ajustes e otimizações contínuos, mas também destaca a relevância de avanços constantes na capacidade da IA para assegurar resultados mais robustos e generalizados.

6 CONCLUSÃO

A abordagem codificador-decodificador demonstra, por meio dos resultados obtidos, a significativa relevância e eficácia desta proposta para o turismo inclusivo. A modularidade

inerente ao modelo destaca sua flexibilidade e adaptabilidade, fundamentais para atender às demandas específicas.

Além de contribuir para a pesquisa em tecnologia assistiva, este trabalho estabelece uma base sólida para a implementação efetiva de soluções inovadoras, promovendo ativamente a inclusão no setor turístico. Ao reconhecer e abordar as barreiras enfrentadas pela comunidade de pessoas com deficiências visuais, o projeto representa um avanço significativo na construção de um turismo mais aberto e acessível.

6.1 Trabalhos Futuros

Na busca por avanços contínuos, uma linha de pesquisa promissora é a expansão do conjunto de dados, incluindo uma variedade mais ampla de contextos turísticos e elementos visuais. Essa abordagem pode aprimorar a capacidade de generalização do modelo, proporcionando um desempenho consistente em diversas situações e ambientes.

Para traduzir esses avanços em benefícios tangíveis, é proposto o desenvolvimento de um aplicativo baseado no algoritmo aprimorado. Este aplicativo teria como foco proporcionar uma experiência de turismo inclusivo, precisamente adaptada a diferentes contextos e elementos visuais encontrados em destinos turísticos. É pretendido colaborar diretamente com destinos turísticos, envolvendo gestores e especialistas em acessibilidade para customizar as funcionalidades do aplicativo de acordo com as características distintas de cada local. A coleta de *insights* valiosos desses *stakeholders* garantirá não apenas a eficácia, mas também a relevância contínua do aplicativo em contextos turísticos específicos.

O envolvimento ativo da comunidade de usuários com deficiência visual será um pilar central no desenvolvimento do aplicativo. Busca-se estabelecer um canal de comunicação aberto para receber *feedback* contínuo, identificar áreas de aprimoramento e implementar melhorias constantes. Acredita-se que essa abordagem colaborativa será favorável para assegurar não apenas a eficácia a curto prazo, mas também a relevância e utilidade a longo prazo do aplicativo.

Adicionalmente, propõe-se o desenvolvimento de uma ferramenta de *plugin* destinada à leitura de imagens. Esta iniciativa visa ampliar as capacidades de processamento de informações visuais, promovendo uma abordagem mais abrangente em relação à análise de conteúdos multimídia. Além disso, planeja-se direcionar estudos para projetos de extensão para o curso de Bacharelado em Ciência da Computação do IFCE - Campus Aracati.

Ao concentrar esforços nessas iniciativas, pretende-se não apenas impulsionar a capacidade do algoritmo em lidar com uma variedade de contextos turísticos, mas também representar um avanço significativo em direção a um turismo verdadeiramente inclusivo, promovendo a acessibilidade e a igualdade de experiências para todos os turistas.

REFERÊNCIAS

- ALZUBAIDI, L. et al. Review of deep learning: Concepts, cnn architectures, challenges, applications, future directions. **Journal of big Data**, Springer, v. 8, p. 1–74, 2021.
- ANACLETO, S. C. **Turismo acessível: criação de um circuito turístico para a população com deficiência visual, na cidade de Braga**. Tese (Doutorado) — Universidade do Minho (Portugal), 2014.
- ANDRADE, A. M. d. S. Os centros urbanos brasileiros e o direito a acessibilidade das pessoas com deficiência. Centro Universitário de Alagoinhas – UNIRB, 2022.
- ARPIT, D. et al. h-detach: Modifying the lstm gradient towards better optimization. **arXiv preprint arXiv:1810.03023**, 2018.
- BIELER, B. D. Acessibilidade como fator de competitividade em hotelaria. Universidade do Estado do Amazonas, 2020.
- CARDOSO, B. R. Inclusão de pessoas com deficiência na publicidade: um estudo sobre a comunicação acessível da marca avon a partir da campanha dona dessa beleza. 2018.
- DUARTE, D. C.; PEREIRA, J. C. R. Acessibilidade para pessoas com deficiência visual: um levantamento preliminar nos setores hoteleiros sul e norte de Brasília. **Cenário: Revista Interdisciplinar em Turismo e Território**, v. 5, n. 9, p. 62–82, 2017.
- DURET, J. et al. Enhancing expressivity transfer in textless speech-to-speech translation. **arXiv preprint arXiv:2310.07279**, 2023.
- FACCIONI, H. Desenvolvimento de uma técnica de ativação de pontos táteis para aplicação em displays braille. 2012.
- FERNANDES, I.; PINHEIRO, H. C. L. Política de acessibilidade universal na sociedade contemporânea. **Textos & Contextos (Porto Alegre)**, 2013.
- FRAGATA, A. R. A. P. et al. Arte-educação e inclusão: Uma reflexão sobre o ensino das artes para deficientes visuais sob uma perspectiva social e educacional. **Cuadernos de Educación y Desarrollo**, Servicios Académicos Intercontinentales SL, n. 113, 2019.
- GRAVES, A. et al. Connectionist temporal classification: labelling unsegmented sequence data with recurrent neural networks. In: **Proceedings of the 23rd international conference on Machine learning**. [S.l.: s.n.], 2006. p. 369–376.
- GROSSBERG, S. Recurrent neural networks. **Scholarpedia**, v. 8, n. 2, p. 1888, 2013.
- HERNÁNDEZ, A.; AMIGÓ, J. M. Attention mechanisms and their applications to complex systems. **Entropy**, MDPI, v. 23, n. 3, p. 283, 2021.
- JARRAHI, M. H. Artificial intelligence and the future of work: Human-ai symbiosis in organizational decision making. **Business horizons**, Elsevier, v. 61, n. 4, p. 577–586, 2018.
- KINGHORN, P.; ZHANG, L.; SHAO, L. A hierarchical and regional deep learning architecture for image description generation. **Pattern Recognition Letters**, v. 119, p. 77–85, 2019. ISSN 0167-8655. Deep Learning for Pattern Recognition. Disponível em: <<https://www.sciencedirect.com/science/article/pii/S0167865517303240>>.

KITA, D. M. Iara app e o paradigma da inclusão: desafios no desenvolvimento de um aplicativo acessível. Universidade Estadual Paulista (Unesp), 2022.

LIMA, J. R. et al. Reconhecimento de voz para inclusão de deficientes visuais em ambientes virtuais de aprendizagem. **Nuevas Ideas en Inform'atica Educativa TISE**, p. 23–29, 2015.

LOCH, R. E. N. Cartografia tátil: mapas para deficientes visuais. **Portal de Cartografia das Geociências**, v. 1, n. 1, p. 36–58, 2008.

LOPES, L. V. C. d. F. et al. Convenção sobre os direitos das pessoas com deficiência da onu, seu protocolo facultativo e a acessibilidade. Pontifícia Universidade Católica de São Paulo, 2009.

MATTOSO, V. d. A. et al. Ora, direis, ouvir imagens?: Um olhar sobre o potencial informativo da áudio-descrição aplicada a obras de artes visuais bidimensionais como representação sonora da informação em arte para pessoas com deficiência visual. Universidade Federal do Rio de Janeiro/Instituto Brasileiro de Informação, 2012.

MELO, A. M.; COSTA, J. d.; SOARES, S. d. M. Tecnologias assistivas. **ACESSIBILIDADE**, p. 62, 2006.

MOURA, D. P. G. Políticas públicas de turismo voltadas para a inclusão social de pessoas com deficiência. 2013.

NASCIMENTO, J. P. d. Qualificação profissional no atendimento da pessoa com deficiência ou mobilidade reduzida nos atrativos turísticos da cidade de manaus. Universidade do Estado do Amazonas, 2018.

NOORD, N. V.; POSTMA, E. Learning scale-variant and scale-invariant features for deep image classification. **Pattern Recognition**, Elsevier, v. 61, p. 583–592, 2017.

O'SHEA, K.; NASH, R. An introduction to convolutional neural networks. **arXiv preprint arXiv:1511.08458**, 2015.

QU, Z. et al. Feedback lstm network based on attention for image description generator. **Computers, Materials & Continua**, v. 59, n. 2, 2019.

REIS, A. C. F. **Economia da cultura e desenvolvimento sustentável: o caleidoscópio da cultura**. [S.l.]: Editora Manole Ltda, 2007.

RODRIGUES, I. M.; PERINOTTO, A. R. C. Comunicação turística acessível a pessoas com deficiências: uma revisão bibliométrica e integrativa da literatura. **Revista Turismo em Análise**, v. 33, n. 2, p. 213–234, 2022.

SILVA, C. A.; FREIRE, A. P. Inspeção da acessibilidade de aplicativos móveis utilizando software leitor de telas. In: SBC. **Anais Estendidos do XVII Simpósio Brasileiro sobre Fatores Humanos em Sistemas Computacionais**. [S.l.], 2018.

SILVA, J. P. D. **O uso das Tecnologias de Informação e Comunicação aplicadas como Tecnologia Assistiva na construção do conhecimento dos alunos com deficiência visual que frequentam as Salas de Recursos Multifuncionais**. Dissertação (Mestrado) — Mestrado em Educação, 2015. Ciências Humanas. Disponível em: <<http://bdtd.unoeste.br:8080/tede/handle/tede/929>>.

SILVA, M. et al. Análise comparativa de sistemas de leitura de tela no apoio ao ensino de pessoas com deficiência visual. In: **Anais do XXVII Workshop de Informática na Escola**. Porto Alegre, RS, Brasil: SBC, 2021. p. 236–244. ISSN 0000-0000. Disponível em: <<https://sol.sbc.org.br/index.php/wie/article/view/17851>>.

STAUEMEYER, R. C.; MORRIS, E. R. Understanding lstm—a tutorial into long short-term memory recurrent neural networks. **arXiv preprint arXiv:1909.09586**, 2019.

STRINGARI, V. Portais governamentais acessíveis: um estudo sobre acessibilidade no portal de saúde. 2021.

SURESH, K. R.; JARAPALA, A.; SUDEEP, P. V. Image captioning encoder–decoder models using cnn-rnn architectures: A comparative study. **Circuits Syst. Signal Process.**, Birkhauser Boston Inc., USA, v. 41, n. 10, p. 5719–5742, oct 2022. ISSN 0278-081X. Disponível em: <<https://doi.org/10.1007/s00034-022-02050-2>>.

TANTI, M.; GATT, A.; CAMILLERI, K. P. Where to put the image in an image caption generator. **Natural Language Engineering**, Cambridge University Press (CUP), v. 24, n. 3, p. 467–489, abr. 2018. ISSN 1469-8110. Disponível em: <<http://dx.doi.org/10.1017/S1351324918000098>>.

TSAIH, R.-H.; HSU, C. C. Artificial intelligence in smart tourism: A conceptual framework. 2018.

TSOTSOS, J. K. **A computational perspective on visual attention**. [S.l.]: MIT Press, 2021.

VALVERDE, J.; OLIVEIRA, A. R. L. de. Além das cortinas visuais: A relação entre tecnologia e acessibilidade na arquitetura de museus. **Revista Projetar-Projeto e Percepção do Ambiente**, v. 7, n. 3, p. 106–116, 2022.

WEERAKODY, P. B. et al. A review of irregular time series data handling with gated recurrent neural networks. **Neurocomputing**, Elsevier, v. 441, p. 161–178, 2021.

WOHFART, E. A.; ROQUE, A. S.; SANTOS, C. P. Navegação indoor baseada em etiquetas rfid: Possibilidades de orientação para deficientes visuais. In: **Proceedings of the 14th Brazilian Symposium on Human Factors in Computing Systems**. [S.l.: s.n.], 2015.

WU, Y. et al. Google’s neural machine translation system: Bridging the gap between human and machine translation. **arXiv preprint arXiv:1609.08144**, 2016.