



UNIVERSIDADE FEDERAL DO PARÁ
INSTITUTO DE CIÊNCIAS EXATAS E NATURAIS
FACULDADE DE COMPUTAÇÃO

Rafael Nascimento de Sousa

**Aplicação de Rede Neural Convolucional na
Classificação de Azulejos Históricos da Cidade
de Belém do Pará**

Belém

2021

Rafael Nascimento de Sousa

Aplicação de Rede Neural Convolucional na Classificação de Azulejos Históricos da Cidade de Belém do Pará

Trabalho de Conclusão de Curso apresentado
na Faculdade de Computação do Instituto de
Ciências Exatas e Naturais como requisito
parcial para obtenção do grau de Bacharel
em Ciência da Computação.

Universidade Federal do Pará

Orientador: Carlos Gustavo Resque dos Santos

Coorientador: Tiago Davi Oliveira de Araújo

Belém

2021

Rafael Nascimento de Sousa

Aplicação de Rede Neural Convolucional na Classificação de Azulejos Históricos da Cidade de Belém do Pará

Trabalho de Conclusão de Curso apresentado
na Faculdade de Computação do Instituto de
Ciências Exatas e Naturais como requisito
parcial para obtenção do grau de Bacharel
em Ciência da Computação.

Conceito: _____

Belém, 9 de abril de 2021.

BANCA EXAMINADORA

Dr. Carlos Gustavo Resque dos Santos - Orientador
UFPA

MsC. Tiago Davi Oliveira de Araújo - Coorientador
UFPA

Dr. Bianchi Serique Meiguins - Membro interno
UFPA

Dr. Anderson Gregório Marques Soares - Membro externo
UFRA

Agradecimentos

Agradeço a Deus, primeiramente, em que eu sempre posso contar.

Gostaria de agradecer a minha família: Minha mãe Maracy e meu pai Marcelo, que sempre me apoiaram e motivaram.

Ao orientador Dr. Carlos Gustavo Resque dos Santos, pela grande ajuda, pela paciência e por não ter me deixado desistir.

Ao Coorientador MsC. Tiago Davi Oliveira de Araújo, por sempre me ajudar e não ter me deixado desistir.

Ao Laboratório - LABVIS, por proporcionarem um ambiente acolhedor, durante os anos de faculdade.

Muito obrigado a todos vocês, por me ajudarem a concluir mais essa etapa da minha vida. Obrigado.

Resumo

A herança cultural de uma cidade é de grande importância para a manutenção e valorização da sua história. Tecnologias inovadoras como Realidade Aumentada e Visão Computacional podem ser utilizadas para dar ênfase a essas heranças que a própria cidade carrega de forma atrativa e lúdica. Na cidade de Belém do Pará, os azulejos são um exemplo importante da herança cultural presente na cidade, que remonta momentos e características de sua fundação. O reconhecimento de imagens, por exemplo, pode facilitar a busca por informações históricas sobre um determinado azulejo de maneira mais rápida, uma vez que com apenas uma foto do azulejo suas informações seriam retornadas ao usuário, dessa forma facilitando a busca e o acesso a informações sobre os azulejos, além da identificação dos mesmos. Tendo em vista o que foi exposto, este Trabalho de Conclusão de Curso apresenta um protótipo que utiliza Redes Neurais Convolucionais (CNN) para classificar, através de imagens, os azulejos históricos de Belém. O treinamento da CNN utilizou três bases de imagens, sendo a primeira composta por imagens que continham azulejos e outros elementos do ambiente (paredes, portas, fachadas dos prédios, etc.), a segunda composta por imagens com foco nos padrões de azulejos e a terceira composta por fotos tiradas da tela do computador, a partir de imagens da segunda base, sendo que uma vez criadas as bases são combinadas em uma base final utilizada no treinamento do modelo. Foram utilizados 12 tipos diferentes de azulejos que representam as classes a serem reconhecidas pela CNN. Dessa forma, após o treinamento é gerado o modelo do tipo tflite (Tensorflow Lite) com as classes de azulejos, para ser utilizado na tarefa de classificação de imagem em dispositivo móvel. Com o modelo e as classes definidas, foi criado para este trabalho de conclusão de curso uma aplicação em que o usuário tira uma foto de um azulejo e são retornadas as quatro classes com as melhores acurácia e as informações históricas dos azulejos classificados. Para a tarefa de classificação de imagens, foi criado um módulo que reúne as funções necessárias para a tarefa de classificação de imagens e que pode ser reutilizado em outros projetos.

Palavras-chave: Reconhecimento de Imagem. Herança Cultural. Azulejo. Redes Neurais Convolucionais. Tensorflow lite.

Abstract

The cultural heritage of a city is of great importance for the maintenance and enhancement of its history. Innovative technologies such as Augmented Reality and Computer Vision can be used to emphasize these legacies that the city itself carries in an attractive and playful way. In the city of Belém do Pará, tiles are an important example of the cultural heritage present in the city, which goes back to the moments and characteristics of its foundation. The recognition of images, for example, can facilitate the search for historical information about a certain tile in a faster way, since with just one photo of the tile, your information would be returned to the user, thus facilitating the search and access to information on the tiles, in addition to their identification. In view of what has been exposed, this Course Conclusion Paper presents a prototype that uses Convolutional Neural Networks (CNN) to classify, through images, the historical tiles of Belém. CNN's training used three image bases, being the first composed of images that contained tiles and other elements of the environment (walls, doors, building facades, etc.), the second composed of images with a focus on tile patterns and the third composed of photos taken from the computer screen, from the second base images, and once created the bases are combined into a final base used in training the model. Twelve different types of tiles were used, representing the classes to be recognized by CNN. Thus, after training, the tflite model (Tensorflow Lite) with the classes of tiles is generated, to be used in the image classification task on a mobile device. With the model and the classes defined, an application was created for this conclusion work in which the user takes a picture of a tile and the four classes are returned with the best accuracy and the historical information of the classified tiles. For the image classification task, a module was created that gathers the necessary functions for the image classification task and that can be reused in other projects.

Keywords: Image recognition. Cultural heritage. Azulejo. Convolutional Neural Networks. Tensorflow lite.

Lista de ilustrações

Figura 1.	Representação padrão de uma Rede Neural Simples. Uma camada de entrada, e uma de saída estão sempre presentes, e várias camadas ocultas (internas) podem constituir a arquitetura de uma Rede Neural.	15
Figura 2.	Arquitetura LeNet, primeira CNN publicada. As indicações na imagem indicam Convoluçãoes, Pooling e Camadas conectadas ao final. Fonte: (LeCun et al., 1998)	16
Figura 3.	Recorte de camadas de uma CNN, exemplificando a variação entre camadas de convolução e <i>pooling</i>	16
Figura 4.	Exemplos de imagens presentes na 1º base de dados (contendo azulejos e outros elementos) e na 2º base de dados (contendo apenas azulejos) respectivamente.	24
Figura 5.	Imagen 2x2 dos doze tipos de azulejos presentes na base de dados. . .	25
Figura 6.	Compilado demonstrativo de imagens de mesma classe, após a padronização das dimensões	26
Figura 7.	Exemplo das transformações aplicadas pelo Data Augmentation.	28
Figura 8.	Matriz de confusão dos resultados de classificação das imagens de azulejos de validação.	30
Figura 9.	Telas da aplicação em ordem de interação. (A) Tela inicial, possui um botão que abre a câmera do dispositivo, para tirar a foto. (B) Tela contendo a foto tirada e o nome das classes juntamente com a acurácia da tarefa de classificação. (C) Tela com detalhes sobre os azulejos, acessada quando o usuário clica em um item listado na tela (B).	31
Figura 10.	Diagrama de Classe UML	33
Figura 11.	Diagrama de Sequência UML	34
Figura 12.	Importando módulo em novo projeto.	36
Figura 13.	Dependências necessárias no build.gradle do novo projeto.	36
Figura 14.	Adicionar modelo (model.tflite) e arquivo com classes (labels.txt) no projeto.	36

Lista de tabelas

Tabela 1.	Tabela com assuntos presentes nos Trabalhos relacionados. Na primeira linha é adicionado o trabalho proposto para melhor comparação.	20
Tabela 2.	Atributos da base sobre as tipologias de Padrões de Azulejos	22
Tabela 3.	Atributos da base sobre as tipologias de guarnições: cercaduras e frisos	22
Tabela 4.	Atributos da base sobre registros devocionais	23
Tabela 5.	Atributos da base sobre painéis e elementos ornamentados	23
Tabela 6.	Sumário da arquitetura selecionada.	28
Tabela 7.	Resultados dos conjunto de dados de treinamento e teste.	29

Lista de abreviaturas e siglas

AM	Aprendizado de Máquina
AP	Aprendizado Profundo
API	<i>Application Programming Interface</i>
BOVW	<i>Bag of Visual Words</i>
CNN	Rede Neural Convolucional — <i>Convolutional Neural Network</i>
CSS	<i>Cascading Style Sheets</i>
CSV	<i>Comma-Separated Values</i>
GPS	<i>Global Positioning System</i>
HTML	<i>HyperText Markup Language</i>
IPHAN	Instituto do Patrimônio Histórico e Artístico Nacional
RA	Realidade Aumentada
RAM	Realidade Aumentada Móvel
UML	<i>Unified Modeling Language</i>
VC	Visão Computacional

Sumário

1	INTRODUÇÃO	10
1.1	Justificativa	12
1.2	Objetivos	12
1.3	Organização do Trabalho	13
2	TEORIA RELACIONADA	14
2.1	Aprendizado de Máquina	14
2.2	Redes Neurais	14
2.3	Redes Neurais Convolucionais	15
3	TRABALHOS RELACIONADOS	18
4	BASE DE DADOS	21
4.1	Base de Informações sobre Azulejos	21
4.2	Base Imagens de Azulejos para Treinamento da CNN	23
5	TREINAMENTO DO MODELO	26
5.1	Pré-Processamento	26
5.2	Seleção da CNN	27
5.3	Treinamento	27
6	IMPLEMENTAÇÃO	31
6.1	Funcionalidades do Protótipo	31
6.2	Modularização	32
6.3	Diagrama de Classes	33
6.4	Diagrama de Sequência	34
6.5	Utilização do Módulo	35
7	CONCLUSÃO	37
	REFERÊNCIAS	39

1 Introdução

Uma das atividades importantes para economia e identidade cultural é o turismo (Awang, Hassan e Zahari, 2009) e dentre as suas várias modalidades, uma das que mais expressa a identidade de uma cidade é a visitação e apresentação da sua herança cultural, tanto em forma física (como edificações, avenidas, *etc.*) como a exposição de suas histórias e costumes associados (Williams, 2014).

A cidade de Belém do Pará (Brasil) é detentora de uma vasta herança cultural, possuindo, por exemplo, patrimônios físicos construídos em diversos momentos da história da cidade. entre os patrimônios físicos, destaca-se o acervo de azulejos fabricados na Europa que foram trazidos para decorar a casa de pessoas influentes entre os séculos XVIII e XX (Arruda e Sanjad, 2017). A quantidade expressiva de azulejos no patrimônio de Belém foi crescente principalmente no período de crescimento econômico da cidade (Alcântara, Brito e Sanjad, 2016), quando ela deixa de ser uma economia de subsistência e passa a ser voltada para o mercado.

Belém do Pará possui uma diversidade expressiva de padrões diferentes de azulejos, sendo que eles foram catalogados em forma de livro (Alcântara, Brito e Sanjad, 2016) publicado pelo IPHAN (Instituto do Patrimônio Histórico e Artístico Nacional) sendo o maior acervo de azulejos da cidade já publicado, contendo vários tipos de azulejos como padronagem, painéis entre outros.

Toda essa riqueza histórica já é apresentada ao turista de maneira narrada ou por meio de livros, mas poderia ter o auxílio de novas tecnologias que transmitissem essa herança cultural de maneira mais interativa e lúdica. Além disso, a própria atividade turística tem se caracterizado cada vez mais pela independência do visitante com o auxílio de tecnologias como aplicativos móveis que auxiliam a visitação e exploração de pontos de interesse e conteúdos associados (Kounavis, Kasimati e Zamani, 2012).

Dentre as áreas de tecnologia da informação que podem ser utilizadas para auxílio do visitante (Godoy et al., 2019) destacam-se as aplicações móveis, cidades inteligentes, redes de computadores, processamento de imagens e visão computacional. Dentro do escopo deste trabalho está a aplicação de técnicas de visão computacional para facilitar a obtenção de conteúdos sobre um determinado azulejo somente a partir de uma foto dele, que pode ser obtida através da câmera de um celular.

Visão Computacional (VC) (Neves, Neto e Gonzaga, 2012) é uma área que está envolvida com a computação gráfica, contém intersecções com áreas como síntese e processamento de imagem além de utilizar muitas técnicas e algoritmos que advêm do Aprendizado de Máquina (AM). A VC visa a aproximação do processo de reconhecimento

feito pelo ser humano e pela máquina (Reis, Albuquerque e Castro, 2001).

Com a constante evolução da tecnologia e a busca crescente por inovação, há uma grande variedade de aplicações para a área de VC, tendo como exemplos de aplicações na área, a ajuda no diagnóstico de doenças (Silva et al., 2020), o reconhecimento de rostos visando aplicação de filtros em aplicativos (Lucena, 2019), além de aplicações que envolvem a questão cultural, com o intuito de apresentar informações adicionais ao reconhecer dada imagem, como em museus e para a área de turismo (Martins, Malta e Costa, 2015).

Dentre alguns dos possíveis problemas ligados a identificação de imagens estão a mudança de iluminação, deformação por propriedades físicas das câmeras, oclusão da cena ou objeto de interesse, mudanças de ponto de vista e “background clutter” (nome técnico para um grande volume de informação que não contribuem para a tarefa desejada) (Zendel et al., 2017) (Pankanti et al., 2011). Dessa forma, com a possibilidade de ocorrência desses problemas, é necessário um método capaz de fazer análises mais precisas, considerando a variabilidade de imagens.

Em geral, entre os modelos mais comuns de AM voltados à classificação (Alpaydin, 2010), tomando como exemplo a detecção de algum objeto em uma imagem, as principais características que definem o objeto devem ser extraídas utilizando técnicas e ferramentas específicas para extração, tornando possível o treinamento e a realização da detecção do objeto nas imagens. Entretanto, esse tipo de abordagem requer encontrar uma técnica de extração de características eficaz, o que demanda bastante tempo e nem sempre alcança um resultado suficientemente confiável. Visando contornar esses problemas, diversas abordagens foram desenvolvidas, dentre elas o Aprendizado Profundo (AP).

Aprendizado Profundo (Indolia et al., 2018) é um tipo de AM, e uma de suas características é ser um modelo de aprendizado que possui diversas camadas, onde cada camada recebe como entrada a saída da camada anterior, afunilando cada vez mais a representação do objeto desejado.

Considerando o que foi exposto, esse trabalho se utiliza de uma das técnicas de AP, a Rede Neural Convolucional (CNN - *Convolutional Neural Network*). A CNN (O’Shea e Nash, 2015) faz a combinação de partes das imagens fornecidas, fazendo comparações entre estas pequenas partes chamadas de máscaras, que são padrões usados para verificar as novas imagens. O conjunto de máscaras formam um filtro, e uma operação de convolução é realizada em cada camada onde há filtros para revelar novos padrões. Os valores para as máscaras nos filtros são aprendidos durante a etapa de treinamento da CNN.

Uma CNN é composta por duas principais camadas: a camada de convolução, que são combinadas inúmeras vezes, gerando assim novas imagens com mapas de ativação do resultado da similaridade, onde uma função de ativação realiza a normalização dos valores dessas ativações obtidas pela convolução; e a camada de subamostragem, comumente

designada de *pooling*, onde a dimensão ou resolução da pilha de imagens fornecidas são reduzidas a uma pilha de imagens de resoluções menores (Haykin, 2008).

Visando a classificação dos azulejos, foram utilizadas duas bases de dados previamente construídas em (Santos et al., 2017), a primeira contendo apenas azulejos e a segunda azulejos e outros elementos adjacentes (como ruas, paredes, vegetação, etc.) na imagem. Dessa forma, foi feito o treinamento do modelo inteligente utilizando essas bases de dados com a finalidade desse modelo classificar azulejos a partir de novas fotos obtidas por usuário da aplicação.

1.1 Justificativa

O turismo, e suas atividades relacionadas, são importantes para uma cidade tanto a nível econômico quanto a nível de preservação e divulgação do seu patrimônio histórico e cultural. Inovações agregadas à essa importante atividade humana tem relevância tanto pelo fato de adaptar essas atividades à evolução tecnológica atual, quanto de promover e dar um envolvimento mais lúdico para a herança cultural de uma cidade.

A cidade de Belém, Pará, possui uma grande riqueza e diversidade de heranças culturais, dentre elas os azulejos visto que a cidade possui um grande acervo. Dessa forma, em virtude da dificuldade de identificar e buscar por informações sobre esses azulejos, este trabalho de conclusão de curso, apresenta uma aplicação que utiliza classificação de imagens, uma das técnicas que estão no estado da arte de VC, agregado com o conteúdo sobre azulejos presentes da cidade de Belém, facilitando o acesso as informações dessa herança cultural.

Em adição, foi criado um módulo que reuni as funções necessárias para a tarefa de classificação de imagens, para que através da utilização de CNNs seja possível classificar azulejos históricos da cidade de Belém do Pará. Em suma, buscando o aprendizado da área de Reconhecimento de imagens e procurando solucionar problemas existentes na área.

1.2 Objetivos

O objetivo do trabalho é utilizar uma Rede Neural Convolucional para Classificar Azulejos Históricos de Belém.

Os objetivos específicos são:

- Criar uma base sobre informações de azulejos históricos em formato digital contendo as informações e imagens dos azulejos;

- Treinar e disponibilizar um modelo de CNN que classifica imagens de azulejos conforme seu tipo;
- Desenvolver módulo classificação de imagem contendo atributos e funções necessárias para realizar a tarefa de classificação, além de levar em conta a facilidade de uso de suas funções pelo programador e a reutilização do módulo em outras aplicações android, que necessitem de um classificador de imagens;
- Desenvolver aplicação que utiliza o módulo criado para a classificação de imagens de azulejos capturados pela câmera de um dispositivo móvel, além de apresentar suas informações históricas.

1.3 Organização do Trabalho

Este Trabalho está dividido em sete capítulos: Introdução, Teoria Relacionada, Trabalhos Relacionados, Base de Dados, Treinamento do Modelo, Implementação e Conclusão.

- **Introdução:** Apresenta de forma geral, as áreas de conhecimento utilizadas e necessárias para a criação do trabalho, além da sua justificativa e objetivos;
- **Teoria Relacionada:** Apresenta os conceitos base para a compreensão do texto;
- **Trabalhos Relacionados:** Apresenta os trabalhos relacionados a este publicados na literatura acadêmica;
- **Base de Dados:** Explica a criação da base de dados disponibilizada e usada no trabalho.
- **Treinamento do Modelo:** Mostra os procedimentos feitos para o treinamento da CNN, assim como, os resultados de eficiência do treinamento.
- **Implementação:** Apresenta do módulo de classificação e a aplicação de reconhecimento de azulejos históricos de Belém para dispositivos móveis.
- **Conclusão:** Conclui o trabalho e define direções para trabalhos futuros.

2 Teoria Relacionada

Este capítulo apresenta conceitos que serão utilizados no trabalho e importantes para o entendimento da pesquisa.

2.1 Aprendizado de Máquina

Com a população tendo acesso a informatização, diversas tecnologias passaram a fazer parte do dia a dia da população de uma forma natural, sendo muitas vezes dificilmente perceptível. O Aprendizado de Máquina (Mohri, 2018) não é uma exceção, sendo caracterizado como um sistema que utilizando suas próprias experiências consegue modificar o seu comportamento, ou seja, suas regras e lógicas necessárias para a execução de uma determinada tarefa, com o mínimo de interferência humana. Essas regras, são determinadas com base no reconhecimento de padrões dentro dos dados analisados.

Existem diversas abordagens para o aprendizado de máquina, sendo o aprendizado profundo ou *deep learning* uma das mais conhecidas. Segundo Liu et al. (2017), o aprendizado profundo é bastante utilizado para solucionar problemas com grandes quantidades de dados, sendo que entre algumas de suas abordagens que obtiveram sucesso, para este trabalho vale destacar as que envolvem visão computacional e reconhecimento de padrões.

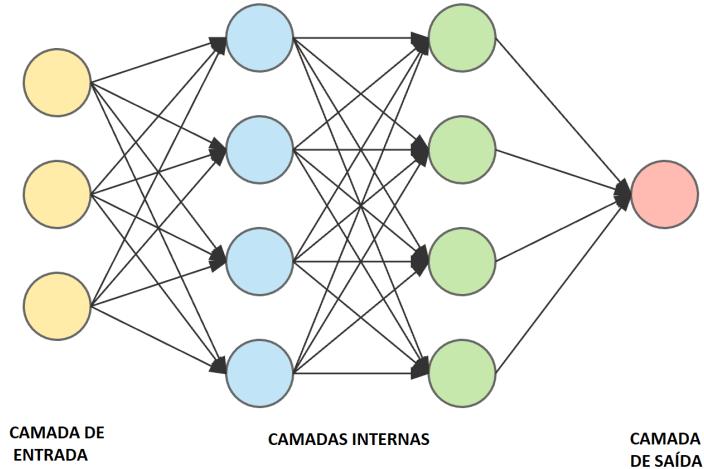
2.2 Redes Neurais

Tecnologias que auxiliam na resolução de problemas são indispensáveis para melhorar questões de eficiência e resolução de tarefas que demandam muito esforço. Assim, um sistema computacional capaz de resolver problemas complexos como o ser humano, seria uma ferramenta de grande ajuda. Em 1943, Warren McCulloch e Walter Pitts produziram a ideia de rede neural, usando como analogia células nervosas e processamento eletrônico, (Rojas, 1996).

Uma Rede Neural (O'Shea e Nash, 2015), é um conjunto de estruturas (ver Figura 1) chamadas de nós, semelhantes a neurônios, conectadas entre si e agrupadas em camadas de entrada, intermediárias (ocultas) e de saída, que possibilitam a identificação de padrões, classificação e agrupamento de dados. Essa estrutura tem a capacidade de aprender padrões, através de ciclos de treinamentos, possibilitando a melhoria de sua acurácia a cada ciclo.

Redes Neurais apresentam excelente resultados em tarefas de cognição humana baseadas em imagem e som (LeCun, Bengio e Hinton, 2015), já que os ciclos de treinamento fazem com que a rede aprenda que características são importantes dentro da tarefa. Em

Figura 1. Representação padrão de uma Rede Neural Simples. Uma camada de entrada, e uma de saída estão sempre presentes, e várias camadas ocultas (internas) podem constituir a arquitetura de uma Rede Neural.



Fonte: retirada de (Costa, 2018)

questão de imagem, fatores como iluminação, deformação, oclusão e objetos semelhantes são considerados menos importantes e não afetam tanto o resultado.

2.3 Redes Neurais Convolucionais

As técnicas convencionais de aprendizado de máquina (LeCun, Bengio e Hinton, 2015), tinham em comum a necessidade de uma complexa e cuidadosa engenharia quanto a extração das características a partir dos dados fornecidos, necessitando até mesmo de pessoas capacitadas e com profundo conhecimento em relação ao domínio do problema, para o auxílio da identificação das características relevantes.

Uma Rede Neural Convolucional (CNN) funciona a base de aprendizado de representações (LeCun, Bengio e Hinton, 2015), onde não é preciso extrair as características já que a própria rede aprende apenas com os dados fornecidos decidindo o que é importante para o problema, cortando assim a necessidade de ter um especialista do domínio e facilitando a extração das características presentes nos dados. Com isso, uma CNN é um modelo matemático que é utilizado em visão computacional (VC), para a resolução de problemas ligados à classificação de imagens. A figura 2 apresenta a primeira CNN publicada, a LeNet.

Uma CNN é composta de várias “camadas”, a principal delas a camada de convolução (O’Shea e Nash, 2015), onde são aplicados filtros (um conjunto de modificadores) em recortes da imagem, buscando a retirada de características. Esses filtros são capazes

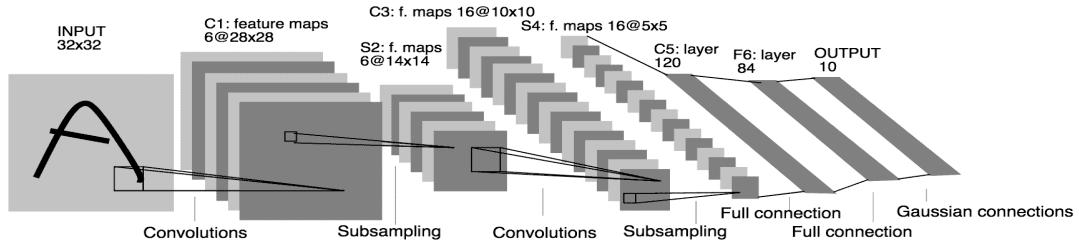


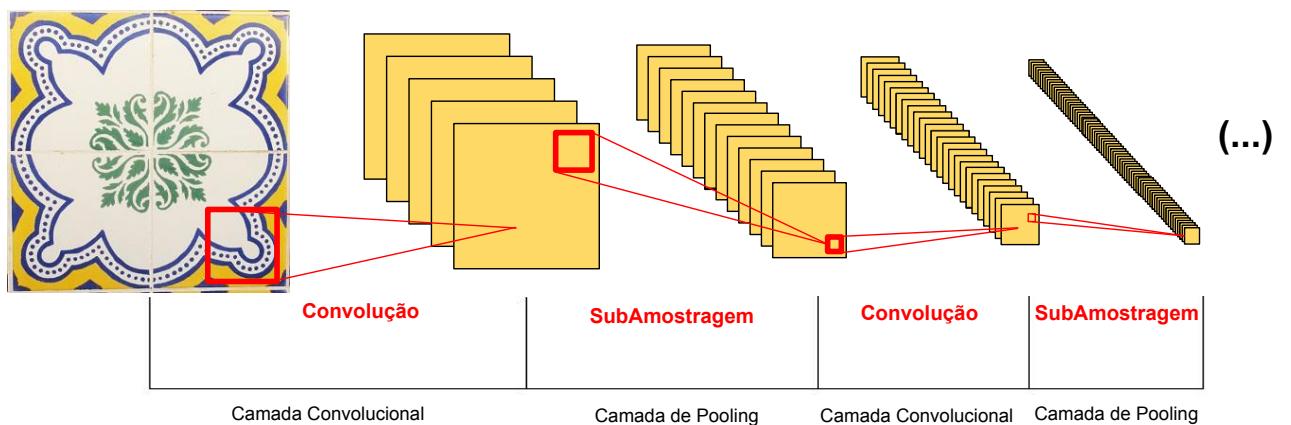
Figura 2. Arquitetura LeNet, primeira CNN publicada. As indicações na imagem indicam Convoluções, Pooling e Camadas conectadas ao final. Fonte: (LeCun et al., 1998)

de retirar características mais complexas camada a camada com maior precisão, sempre levando em conta a melhor forma possível de representar as características presentes na imagem.

A camada de *pooling* (O'Shea e Nash, 2015), é responsável por agrupar os mapas de ativação da imagem e os filtrar em um subconjunto, dessa forma tornando possível, por exemplo, modificar as dimensões da imagem dependendo da situação, mantendo as principais características que foram encontradas, além de diminuir a força de computação necessária. Por fim as camadas totalmente conectadas (O'Shea e Nash, 2015), que produzem pontuações ou votos de classe a partir dos valores de ativações gerados nas camadas anteriores, sendo assim usado para a classificação do padrão.

Dessa forma, a Rede Neural Convolucional necessita de diversas camadas, saindo do reconhecimento de características mais “comuns” para o reconhecimento de características mais complexas, com o passar das etapas, variando entre etapas de convolução e *pooling* (ver Figura 3).

Figura 3. Recorte de camadas de uma CNN, exemplificando a variação entre camadas de convolução e *pooling*



Fonte: criado pelo autor

Após as etapas iniciais, o treinamento é a etapa de aprendizado da rede neural (Zhou, 2018), consistindo de um processo interativo de ajustes sobre os pesos. Esses ciclos são baseados no método do gradiente, um método de otimização de função para encontrar o menor valor de uma função. No caso da CNN, foi feita a otimização da rede para encontrar o menor erro.

Dessa forma, uma rede necessita de muitos ciclos de treino e verificação, para que os pesos possam ser ajustados a cada ciclo, tendendo a diminuir o máximo possível o erro, sendo feitas verificações sobre a discrepância dos valores obtidos em relação aos valores esperados.

3 Trabalhos Relacionados

Neste capítulo, são apresentados os trabalhos relacionados aos principais assuntos abordados nesta pesquisa, apresentando os objetivos, as técnicas utilizadas e os resultados.

Em Rocha (2017), o autor faz uma análise sobre realidade aumentada visando as possibilidades que esta área pode proporcionar para o jornalismo, tendo como objetivo, a inovação da maneira como as notícias são apresentadas e novas maneiras de interação com o leitor. Neste trabalho, foi apresentada a evolução cronológica da tecnologia de realidade aumentada, além dos usos no meio jornalístico utilizando principalmente QRcode, por exemplo, em capas de revistas. Como resultado, a adesão à tecnologia depende do conhecimento do usuário sobre como utilizá-la, dos diferentes possíveis usos e das mudanças na sociedade no futuro.

Em Godoy et al. (2019), os autores apresentam o desenvolvimento de uma aplicação mobile, voltada para o turismo, que permite a exploração e acesso a informações, de âmbito histórico e cultural, em diversos locais pelo município de Santiago/RS. Dos benefícios obtidos, destaca-se a possibilidade de obtenção de conhecimentos culturais e históricos acerca da localidade.

Em Silva et al. (2020), são apresentadas três implementações de redes neurais convolucionais: ResNet50, VGG-16 e InceptionV3, com o objetivo de determinar qual entre as três tem o melhor desempenho na classificação de imagem, para fins médicos. O problema a ser solucionado é a dificuldade de interpretação de imagem de exames, devido à complexidade e importância de um resultado preciso. Dessa forma, as estruturas e etapas de cada implementação foram apresentadas e foi feito um estudo comparativo, utilizando uma base de dados contendo imagens de raio-x de vários tipos de pneumonia (normal, bacterial e viral), de diversas faixas etárias de pacientes.

Foi utilizada a técnica “Data Augmentation” para criar exemplos sintéticos durante o treinamento do modelo, possibilitando a redução do *overfitting*) em redes neurais profundas. Dentro as três arquiteturas testadas, a ResNet50 é a que obteve o melhor desempenho, alcançando 95,10% de f1-score, onde 785 imagens são classificadas corretamente, de um total de 830, análise que é visualizada através da matriz de confusão.

Em Juraszek (2014), os autores têm como objetivo o reconhecimento de imagens de produtos, com base em três métodos: palavras visuais usando descritores artificiais Bag of Visual Words (BOVW), CNNs e descritores naturais (obtidos através de uma rede neural previamente treinada em uma base distinta). Destacando a rede neural convolucional, contendo três camadas de convoluções, que apresentou o melhor resultado entre as técnicas selecionadas, além de tornar efetivo o uso de CNN como um descritor natural (transferência

de conhecimento), onde pesos podem ser reutilizados mesmo em bases distintas.

Em Martins, Malta e Costa (2015), os autores apresentam um estudo sobre realidade aumentada e seu uso para o turismo, este que segundo os autores do trabalho apresenta um grande crescimento de investimentos em tecnologias. Além disso, é apresentado o projeto “Viseu na Palma da Mão”, com o objetivo de usar o recurso de celulares para, de forma geral, que o usuário possa conhecer melhor a cidade. A ferramenta permite, além de visualizar, a manipulação de objetos virtuais apenas apontando o aparelho para o objeto.

Os trabalhos dos autores (He, Wu e Li (2018); Trier, Reksten e Løseth (2021); Lei et al. (2020)) utilizam realidade aumentada por imersão e processamento de imagens, usando como base de dados imagens de museu e arte, a fim de proporcionar experiência e conhecimento de herança cultural.

Os trabalhos de (Marto et al. (2021); Poux et al. (2020); Conde (2020)) apresentam aplicações de realidade virtual por imersão, aprendizado profundo e redes neurais, usando imagens de locais de patrimônio cultural, com a finalidade de proporcionar experiências do visitante em diversos ambientes externos do patrimônio cultural.

Em (Lambers, Vaart e Bourgeois (2019); Trier, Reksten e Løseth (2021); Godewithana et al. (2020); Bonhage et al. (2021)) usam novos métodos de redes neurais profundas, processamento de imagens e a realidade aumentada em imagens arqueológicas, com o propósito de gerar detecções de objetos arqueológicos desconhecidos, de diversos locais, de forma inovadora, precisa e eficiente de diversos locais.

Em Kysela e Štorková (2015) aplicam a tecnologia de realidade aumentada como método de ensino de história e turismo, através de uma aplicação móvel, baseada em geolocalização, com o intuito de oferecer maneiras inovadoras de ensino de forma atraente e eficaz.

Em (Alfonso et al. (2021); Huang (2021); Cauchi e Scerri (2019); Han et al. (2020); Demir e Karaarslan (2018); Fukada, Kasai e Ohtsu (2014); Llerena, Andina e Grijalva (2018)) usam recursos de realidade aumentada, aprendizado profundo e reconhecimento de imagens, por meio de imagens, geolocalização e imersão, com a finalidade de proporcionar passeios turísticos, por diversos pontos da cidade.

A Tabela 1, apresenta a classificação quanto à utilização das tecnologias em relação aos artigos escolhidos para os trabalhos relacionados, além disso, é adicionado na primeira linha o trabalho proposto para melhor comparação com os trabalhos relacionados. Dando enfase as tecnologias de RA, Reconhecimento de imagens e os tipos utilizados. Destaca-se que M&A é uma sigla para Museu e Artes criada para encurtar a tabela.

Tabela 1. Tabela com assuntos presentes nos Trabalhos relacionados. Na primeira linha é adicionado o trabalho proposto para melhor comparação.

Referência	Tec. de RA	Tec. de VC	Área
Trabalho proposto	-	Aprendizado de máquina, CNN	Turismo
Fukada, Kasai e Ohtsu (2014)	GPS	Reconhecimento de imagens por RA	Turismo
Kysela e Štorková (2015)	GPS	-	Ensino
He, Wu e Li (2018)	Imersão	-	M&A
Demir e Karaarslan (2018)	GPS	IFT, SURF	Turismo
Llerena, Andina e Grijalva (2018)	GPS	-	Turismo
Godoy et al. (2019)	GPS	R-CNN	Turismo
Cauchi e Scerri (2019)	GPS	-	Turismo
Lambers, Vaart e Bourgeois (2019)	-	Aprendizado de máquina, CNN	Arqueologia
Poux et al. (2020)	-	Nuvens de pontos, reconstrução 3D	Patrimônio
Godewithana et al. (2020)	Imagen	SURF, CNN, SVM	Arqueologia
Han et al. (2020)	-	TU-DJ-Cluster, aprendizado profundo	Turismo
Hoang et al. (2020)	-	Deep Feature, image Stitching	Turismo
Conde (2020)	Imagen	-	Patrimônio
Lei et al. (2020)	-	VGG	M&A
Huang (2021)	Imagen	-	Turismo
Trier, Reksten e Løseth (2021)	-	R-CNN	Arqueologia
Marto et al. (2021)	Multissensorial	-	Patrimônio
Alfonso et al. (2021)	Modelo 3D	-	Turismo
Bonhage et al. (2021)	-	R-CNN	Arqueologia

Fonte: criado pelo autor

4 Base de Dados

Neste capítulo são apresentadas as bases de dados utilizadas neste trabalho, além das suas características e estrutura e o objetivo para a sua utilização. Foram desenvolvidos tipos de bases, uma para registrar digitalmente conteúdo histórico/cultural acerca dos azulejos e outro para realização do treinamento do algoritmo reconhecedor de azulejos.

4.1 Base de Informações sobre Azulejos

A partir do livro de Alcântara, Brito e Sanjad (2016), com o inventário da azulejaria de Belém do Pará que contém um banco de dados acerca do patrimônio cultural dos azulejos de Belém, foi feita a transcrição das tabelas presentes no livro (que está em formato impresso) para o formato digital, utilizando-se de um editor de planilhas eletrônicas.

Para cada tabela presente do livro, foi criada uma planilha eletrônica contendo uma tabela com as informações separadas em colunas, tendo como principais atributos: código de identificação, endereço, bairro e informações quanto a fabricação, características, aplicações e imagens de azulejos. Após o registro das tabelas na planilha eletrônica, as informações também foram exportadas para o formato CSV (Comma-Separated Values), para que seja facilmente importada em outras aplicações. Dessa forma, o objetivo de transformar essa informação para o formato digital é viabilizar sua utilização por sistemas computacionais, facilitando assim, a criação de aplicações com esse tema. As tabelas, os CSVs e informações retiradas, estão disponíveis online¹.

As imagens de azulejos foram obtidas através de um processo de escaneamento, visto que os dados presentes no livro não estão disponíveis em formato digital. Dessa forma, a Tabela 2 apresenta um total de 274 instâncias de padrões de azulejos, a Tabela 3 apresenta um total de 149 instâncias de cercaduras e frisos, a Figura 4 apresenta um total de 44 instâncias de registros devocionais e a Figura 5 apresenta um total de 12 instâncias de painéis e elementos ornamentados.

A Tabela 2 mostra os oito atributos para descrever cada um dos azulejos de padrões catalogados na base de dados. A primeira coluna mostra o nome dos atributos, a segunda tem uma breve descrição desse atributo e a última coluna mostra um exemplo contido na base de dados.

A Tabela 3 mostra os oito atributos para descrever cada um dos azulejos de guarnições (podem ser entendidos metaforicamente como bordas ou adornos) catalogados na base de dados. A primeira coluna mostra o nome dos atributos, a segunda tem uma

¹ <<http://labvis.ufpa.br/AzulejAR/>>

Tabela 2. Atributos da base sobre as tipologias de Padrões de Azulejos

Nome do Atributo	Descrição	Exemplo
Azulejo-Código	ID atribuído para identificação única dos registros	Azulejo-PE1-1-1
Procedência	Local original de fabricação da peça	Lisboa/Portugal
Fábrica	Nome da Fábrica Original	Viúva Lamego-Constância (Consta do Catalogo de J. Lino de 1889). Apelido: Ferradura
Dimensão-Formato	Tamanho da peça individual de azulejo	13,4 x 13,4cm
Técnica de Fabricação	Material e técnicas utilizadas na para a fabricação do azulejo	Cerâmica e vidrado Estampilha
Endereço	Locais na cidade de Belém onde é possível encontrar tais azulejos	Trav. Frutuso Guimarães n° 18/30/36, 257/259 ...
Bairro	Bairro que contém os endereços	Cidade Velha/Campina
Uso-Aplicação	local e técnica de aplicação do azulejo	Revestimento de fachada, assentados por compressão contra argamassa fresca.

Fonte: versão digital e adaptada de (Alcântara, Brito e Sanjad, 2016)

breve descrição desse atributo e a última coluna mostra um exemplo contido na base de dados.

Tabela 3. Atributos da base sobre as tipologias de guarnições: cercaduras e frisos

Nome do Atributo	Descrição	Exemplo
Azulejo-Código	ID atribuído para identificação única dos registros	FE 15-1-1
Procedência	Local original de fabricação da peça	Lisboa/Portugal
Fábrica	Nome da Fábrica Original	Viúva Lamego
Dimensão-Formato	Tamanho da peça individual de azulejo	13 x 6,3 cm
Técnica de Fabricação	Material e técnicas utilizadas na para a fabricação do azulejo	Cerâmica e vidrado Estampinha
Endereço	Locais na cidade de Belém onde é possível encontrar tais azulejos	Rua Senador Lemos n° 573 ...
Bairro	Bairro que contém os endereços	Campina/Reduto/Umarizal
Uso-Aplicação	local e técnica de aplicação do azulejo	Revestimento de fachada, assentados por compressão contra argamassa fresca.

Fonte: versão digital e adaptada de (Alcântara, Brito e Sanjad, 2016)

A Tabela 4 mostra os cinco atributos para descrever cada um dos registros devocionais (um painel de azulejo dedicado à devoção religiosa) catalogados na base de dados. A primeira coluna mostra o nome dos atributos, a segunda tem uma breve descrição desse atributo e a última coluna mostra um exemplo contido na base de dados.

Tabela 4. Atributos da base sobre registros devocionais

Nome do Atributo	Descrição	Exemplo
Azulejo-Código	ID atribuído para identificação única dos registros	Registro Devocional RD 1-1-1
Técnica de Fabricação	Material e técnicas utilizadas na para a fabricação do azulejo	Estampagem Mecânica
Tema	Tema Devocional	Nossa Senhora de Fátima
Endereço	Locais na cidade de Belém onde é possível encontrar tais azulejos	Travessa Quintino Bocaiúva nº 945
Bairro	Bairro que contém os endereços	Nazaré

Fonte: versão digital e adaptada de (Alcântara, Brito e Sanjad, 2016)

A Tabela 5 mostra os cinco atributos para descrever cada um dos painéis e elementos ornamentados catalogados na base de dados. A primeira coluna mostra o nome dos atributos, a segunda tem uma breve descrição desse atributo e a última coluna mostra um exemplo contido na base de dados.

Tabela 5. Atributos da base sobre painéis e elementos ornamentados

Nome do Atributo	Descrição	Exemplo
Azulejo-Código	ID atribuído para identificação única dos registros	Painel Ornamento PO1
Técnica de Fabricação	Material e técnicas utilizadas na para a fabricação do azulejo	Cerâmica e Vidrado Meio-relevo
Endereço	Locais na cidade de Belém onde é possível encontrar tais azulejos	Travessa Quintino Bocaiúva nº 1455
Bairro	Bairro que contém os endereços	Nazaré
Uso-Aplicação	local e técnica de aplicação do azulejo	Ornato de fachada, assentados por compressão contra argamassa fresca.

Fonte: versão digital e adaptada de (Alcântara, Brito e Sanjad, 2016)

4.2 Base Imagens de Azulejos para Treinamento da CNN

Apesar dos dados e imagens obtidas das tabelas, para criar uma base de dados adequada para a realização do treinamento com a rede neural, é necessário um volume maior de imagens com variedade de tamanhos, iluminação, deformação e distância, para que o resultado final seja confiável.

Foi utilizada uma CNN para a classificação de azulejo, pois o domínio é bem específico sendo melhor usar a CNN que aprende as características sozinha. Com isso à base de dados selecionada (Santos et al., 2017), conta com fotos de azulejos da cidade

tiradas de três aparelhos celulares diferentes. Sendo divididas em duas bases de dados, uma contendo 191 fotos de azulejos e outra 865 imagens com foco nos padrões dos azulejos. Adicionalmente, em decorrência a pandemia de Covid-19 houve a impossibilidade de visitar o interior de certos locais, criando um empecilho na obtenção de novas imagens, foi outra causa que interferiu na seleção da base de dados. A base de dados conta com imagens de mansões históricas da cidade de Belém, com azulejos decorando os interiores e as fachadas. A partir das imagens, Figura 4, foram criadas duas bases de dados, que são usadas efetivamente para o treinamento.

Figura 4. Exemplos de imagens presentes na 1º base de dados (contendo azulejos e outros elementos) e na 2º base de dados (contendo apenas azulejos) respectivamente.



Fonte: (Santos et al., 2017)

A 1º base de dados é composta por 191 imagens, coletadas utilizando três diferentes celulares, sendo que as imagens apresentam diferentes tipos de qualidade, ângulos, sombras, além de outros elementos presentes (portas, janelas e elementos das fachadas dos edifícios), que podem vir a dificultar a identificação dos azulejos presentes nas imagens. Em suma, é a que melhor representa a utilização da aplicação pelo usuário.

A 2º base de dados é composta por 865 imagens, sendo constituída de fragmentos ou pedaços criados a partir das 191 imagens da 1º base de dados. Tem como objetivo, filtrar as imagens com o foco nos padrões dos azulejos e descartando outros elementos presentes nas imagens.

Devido a situação de pandemia do novo corona vírus, período de desenvolvimento deste trabalho de conclusão de curso, e em respeito as normas de segurança sanitária para diminuir o contágio do vírus, o teste do protótipo de forma presencial ficou inviabilizado. Dessa forma, foram adicionadas imagens de azulejos através de fotos da tela do computador. Essa imagens são suscetíveis a interferências na iluminação, distorções, entre outros. Desse modo, são treinadas e testadas imagens, pelo algoritmo, capturadas do local o qual os

azulejos estão e da tela do computador.

A base de dados contém um total de 1164 imagens, distribuídas em doze diferentes classes de azulejo. Para criação do modelo de classificação, a base de dados foi dividida em 80% das imagens para o treinamento e geração do modelo, e 20% para o teste do modelo gerado, sendo 931 e 233 imagens respectivamente.

Com isso, a base de dados gerada pelo treino apresenta doze classes de tipos distintos de azulejos, a Figura 5 mostra os doze tipos de azulejos presentes na base de dados de forma detalhada, apresentados em uma grade 2x6 (altura e largura) para ilustrar as classes do treinamento.

Figura 5. Imagem 2x6 dos doze tipos de azulejos presentes na base de dados.



Fonte: (Santos et al., 2017)

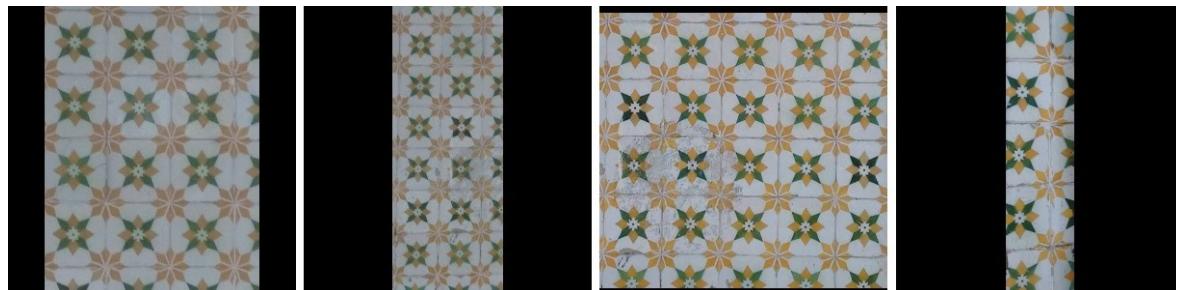
5 Treinamento do Modelo

Neste Capítulo, serão apresentadas etapas referentes ao período de treinamento do modelo, desde o pré-processamento das imagens, a seleção da rede neural convolucional utilizada e o treinamento.

5.1 Pré-Processamento

Nessa etapa, é utilizado um pré-processamento para a padronização das dimensões das imagens presentes nas duas bases de dados. Os dados necessários para o treino são carregados em disco e é feita a padronização das dimensões das imagens presente nas duas bases de dados, adicionando barras pretas verticais ou horizontais e para redimensionar as imagens em de 256×256 pixels sem esticar os azulejos. A figura 6, apresenta um compilado demonstrativo de quatro imagens presentes em uma das doze classes de azulejos, de forma a apresentar como as imagens ficam após o processo de padronização.

Figura 6. Compilado demonstrativo de imagens de mesma classe, após a padronização das dimensões



Fonte: criado pelo autor

Além disso, é importante salientar que é usado o pré-processamento de média zero e desvio padrão unitário para normalizar as imagens de entrada da rede. Aqui, cada amostra ou imagem é normalizada pela subtração da média de pixels de toda a imagem, seguida da divisão pelo desvio padrão das valores de pixels da imagem. Ao final, cada imagem estará com média e desvio padrão dos valores dos pixels próximos de zero e um, respectivamente.

$$I_n = \frac{I - \mu}{\sigma} \quad (5.1)$$

A Equação 5.1 apresenta a normalização realizada por imagem. A imagem antes da normalização é representada por I , e μ e σ são a média e o desvio padrão da imagem

I , enquanto que I_n representa a imagem normalizada cujos valores de pixels na imagem possuem média zero e desvio padrão unitário.

5.2 Seleção da CNN

Foi realizada uma busca aleatória para encontrar a arquitetura e os parâmetros da CNN utilizada na classificação de azulejos, onde dado um conjunto finito de parâmetros, é investigada uma combinação aleatória destes para encontrar os que melhor se ajustam a tarefa proposta neste trabalho (Bergstra e Bengio, 2012).

Foram feitas buscas na quantidade de camadas convolucionais e conectadas, além da quantidade de filtros e neurônios destas, respectivamente. Apesar disso, as arquiteturas predefinidas ResNet50, VGG-19 e InceptionV3 também foram testadas na classificação do azulejos.

Ao finalizar essa busca, observou-se que as arquiteturas predefinidas não apresentaram um bom desempenho na classificação de azulejos, o quais nenhuma desta alcançou uma acurácia superior a 87%. Porém, a busca aleatória para encontrar uma arquitetura própria foi bem sucedida. A arquitetura obtida possui 18 camadas, conforme mostra a Tabela 6.

A arquitetura da CNN para a classificação dos azulejos contém sete camadas convolucionais, sete camadas pooling e quatro camadas conectadas, conforme apresenta a Tabela 6. As camadas convolucionais contém de filtros de tamanho 3×3 e quantidade conforme é apresentado na coluna Filtros/neurônios da Tabela 6. As três primeiras camadas conectadas possuem 1024, 512 e 128 neurônios, respectivamente. Enquanto a quarta e última camada conectada apresenta 12 neurônios, representando as 12 classes de azulejos.

5.3 Treinamento

O otimizador RMSprop é utilizado para minimizar a função de custo ou perda (conhecido como *loss*), com o intuito reduzir a diferença entre a saída esperada e a saída obtida pela CNN (Bishop, 2006). A taxa de aprendizagem utilizada no treinamento foi de 0,001. Quanto as interações, foram usadas 100 épocas e lotes (*batchs*) de 16 imagens.

Considerando que o conjunto de dados de treinamento é relativamente pequeno, foram gerados lotes de imagens sintéticas em tempo real para o treinamento do modelo, através de transformações realizadas pela técnica ‘Data Augmentation’ (Krizhevsky, Sutskever e Hinton, 2012). As transformações aplicadas foram giro horizontal, translação variando o pixels entre -30 e 30 nos eixos vertical e horizontal, além de transformações na escala variando entre 70-130% do tamanho da imagem por eixo.

Tabela 6. Sumário da arquitetura selecionada.

Camada	Tipo	Filtros/neurônios	Número de parâmetros
1	Convolucional	64	1792
2	Pooling		
3	Convolucional	64	36928
4	Pooling		
5	Convolucional	32	18464
6	Pooling		
7	Convolucional	32	9248
8	Pooling		
9	Convolucional	16	4624
10	Pooling		
11	Convolucional	16	2320
12	Pooling		
13	Convolucional	16	2320
14	Pooling		
15	Conectada	1024	66560
16	Conectada	512	524800
17	Conectada	128	65664
18	Conectada	12	1548
Total de parâmetros			734.268
Parâmetros treináveis			734.268

Fonte: criado pelo autor

A Figura 7 apresenta algumas transformações realizadas pelo Data Augmentation. A imagem mais à esquerda representa uma imagem original, sem transformações, de um azulejo da classe soberano, já as imagens subsequentes representam as transformações (incluindo combinações) aplicadas a imagem original.

Figura 7. Exemplo das transformações aplicadas pelo Data Augmentation.



Fonte: criado pelo autor

A Tabela 7 apresenta os resultados obtidos, em termos de acurácia e *loss*, após a geração do modelo para classificação de azulejos. Os valores obtidos pela classificação das

imagens de treinamento mostram que esses dados se ajustam corretamente à aprendizagem do modelo. O resultado de acurácia, 98,60% do conjunto de treinamento indica que quase a totalidade das imagens de treinamento foram classificadas corretamente. Além disso, o *loss*, 0,0622 ver Tabela 7, dos dados de treinamento mostra a minimização eficiente realizada pelo RMSprop, o qual consegue minimizar significativamente as diferenças entre as saídas esperadas e obtidas pela rede.

Tabela 7. Resultados dos conjunto de dados de treinamento e teste.

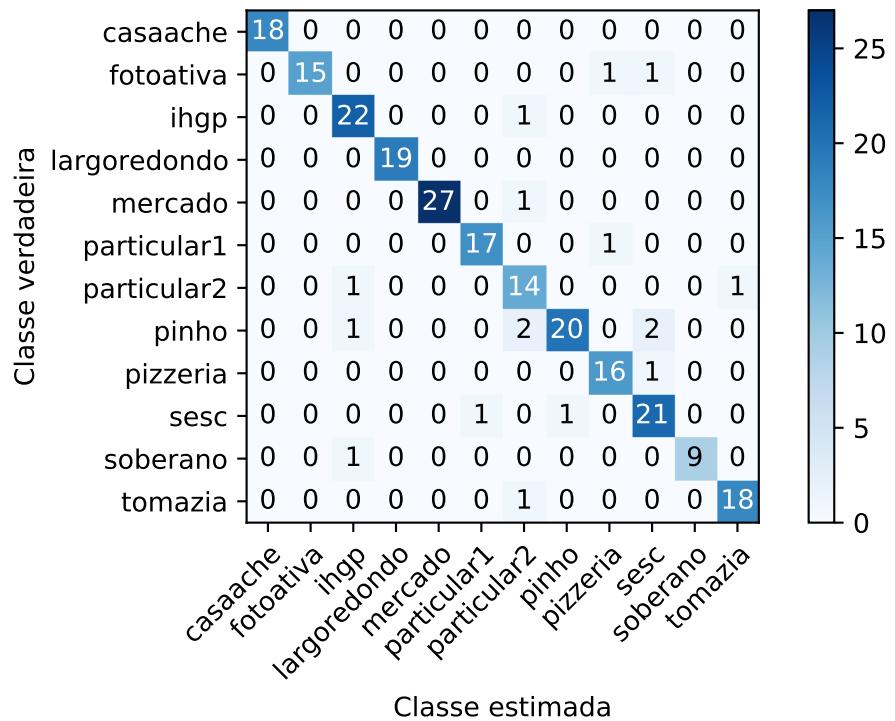
	Acurácia (%)	Loss
Treinamento	98,60	0,0622
Teste	92,70	0,2875

Ao analisar os resultados obtidos pelo conjunto de teste, nota-se a qualidade tanto em termos de acurácia, quanto em termos de *loss*. A acurácia de 92,70%, como mostra a Tabela 7, indica que das 100 imagens utilizadas para o teste do modelo, somente quatro foram assinaladas erroneamente quanto a classe de azulejo. Além disso, o valor 92,70% da acurácia de teste e o valor 98,60% da acurácia de treinamento indicam que não houve sobreajuste (do inglês, *overfitting*) do modelo, ou seja, o modelo classificador de azulejos mostrou-se eficiente ao generalizar e classificas novas imagens, as quais não foram designadas ao treinamento. Por fim, assim como nos resultados de treinamento, o valor de *loss* de teste mostra a qualidade da otimização feita pelo RMSprop.

A base de dados apresenta certo nível de desbalanceamento entre a quantidade de imagens de azulejos por classe. Por isso, para melhor elucidar os resultados obtidos pelas imagens de teste é gerada a matriz de confusão, como mostra a Figura 8. Em classe verdadeira é mostrada a quantidade de imagens pertencem a determinada classe de azulejo, enquanto que classe estimada representa as previsões feitas pelo modelo gerado.

Dezessete imagens são classificadas erroneamente, como exemplifica a Figura 8. Destaca-se a classe “pinho”, cinco imagens das vinte e cinco utilizadas para o treinamento são assinaladas incorretamente, duas são classificadas como sendo os azulejos ‘particular2’ e “sesc”, enquanto uma é classificada como “ihgp”.

Figura 8. Matriz de confusão dos resultados de classificação das imagens de azulejos de validação.



6 Implementação

Neste Capítulo, serão apresentados pontos importantes quanto a implementação do módulo de classificação de imagens, apresentação dos métodos presentes, além do layout da aplicação que utilizou o módulo criado.

6.1 Funcionalidades do Protótipo

Nesta seção são apresentadas as principais telas do protótipo reconhecedor de azulejos. As principais interações do aplicativo desenvolvido são: tirar foto de azulejo, selecionar azulejo correto da lista de possibilidades e visualizar conteúdo sobre o azulejo selecionado.

A Figura 9a apresenta a tela inicial do protótipo, onde a câmera é aberta para capturar a imagem que contenha o azulejo a ser reconhecido. Após a captura da imagem, através do botão FOTO, a imagem capturada é exibida na tela do smartphone.

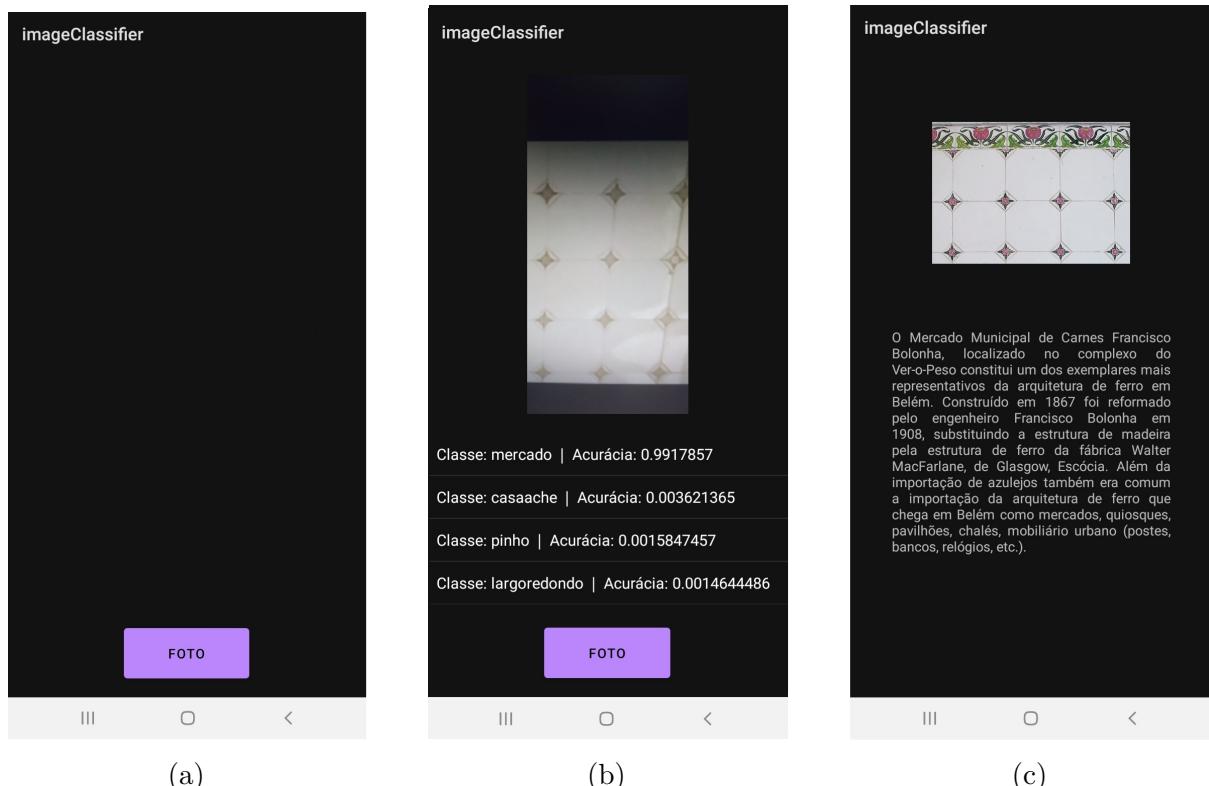


Figura 9. Telas da aplicação em ordem de interação. (A) Tela inicial, possui um botão que abre a câmera do dispositivo, para tirar a foto. (B) Tela contendo a foto tirada e o nome das classes juntamente com a acurácia da tarefa de classificação. (C) Tela com detalhes sobre os azulejos, acessada quando o usuário clica em um item listado na tela (B).

A Figura 9b apresenta a tela de classificação do protótipo. Nela, são apresentadas as quatro classes com maior chance de ser a correta e sua respectiva probabilidade. Os itens da tela são ordenados de cima para baixo em ordem decrescente de probabilidades. No caso mostrado na Figura 9b a lista mostra as classes “mercado” com ~99% de probabilidade, “casaache” com ~0%, “pinho” com ~0% e “largoredondo” com ~0%.

Já a tela apresentada na Figura 9c, mostra o conteúdo referente a classe de azulejo selecionada pelo usuário na lista, sendo que no caso desse print de tela foi a classe “mercado”. Essa tela exibe uma imagem do azulejo juntamente com um texto que contém informações sobre sua história, período de fabricação e localização.

6.2 Modularização

Programação modular, de forma geral, tem como objetivo a separação das “responsabilidade” presentes em determinado programa, visando o agrupamento de um conjunto de variáveis e funções que interagem entre si, dessa forma, constituindo um módulo.

Em termos de estrutura e visibilidade das funções e atributos presentes no módulo, foram utilizadas características de padrões de projeto estrutural *facade* (Gamma et al., 1995), sendo o acesso a complexidade do módulo dividida em pública, onde são definidas as funções que o usuário tem acesso e a parte privada, que contém funções e mecanismos mais complexos, fora de alcance para o usuário, dessa forma, encapsulando a complexidade das funções de classificação de imagem e apresentando ao programador, apenas o essencial para realizar a tarefa de classificação. Dentre as vantagens de utilizar programação modular (Medina, 2006), destaca-se:

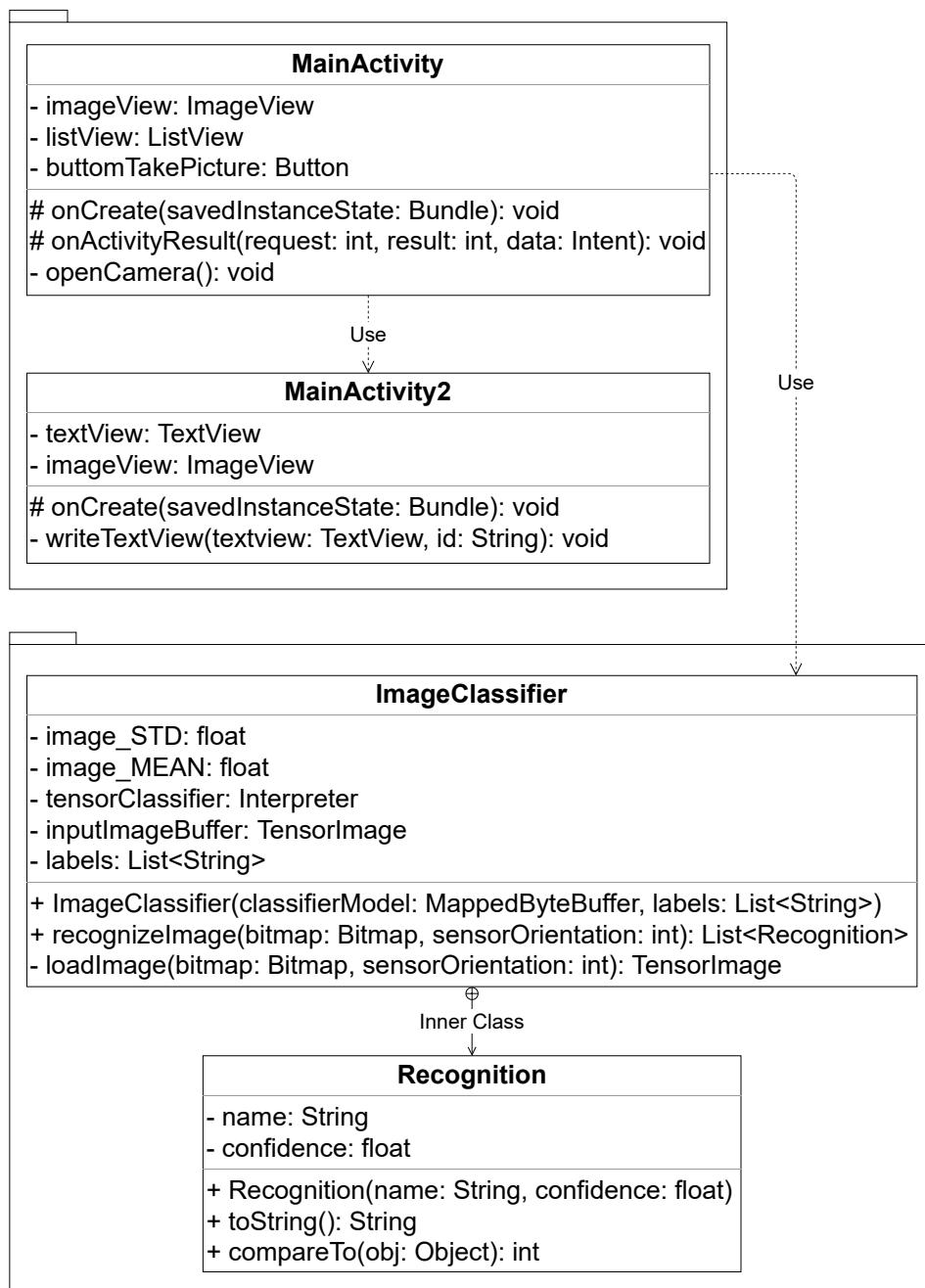
- A facilidade da gerência do desenvolvimento de software, dada a clareza da ferramenta disponível, o que ela oferece e seu objetivo;
- A facilidade no controle e na questão da garantia da qualidade, devido ao “espaço delimitado” dentro do módulo;
- Reutilização do módulo desenvolvido, possibilitando uma maior vida útil.

Para este trabalho, foi criado um módulo responsável pela tarefa de leitura do modelo treinado, e execução da classificação de imagem. O módulo foi criado utilizando um ambiente de desenvolvimento para a plataforma android, Android Studio, com o intuito de reunir as funções e passos necessários para a realização da tarefa de classificação, de forma independente da aplicação. Dessa forma, possibilitando a reutilização e integração do módulo em outros projetos.

6.3 Diagrama de Classes

Com o objetivo de apresentar a estrutura e relação entre as classes do projeto e o módulo de classificação de imagens, a Figura 10 apresenta o diagrama de classe criado a partir do projeto. Nele, foram definidas todas as classes utilizadas, contando com os seus atributos, as suas funções e as relações entre elas.

Figura 10. Diagrama de Classe UML



As classes “MainActivity” e “MainActivity2” são responsáveis pelo controle das interações entre o usuário e a interface, sendo também o papel da “MainActivity”, o envio

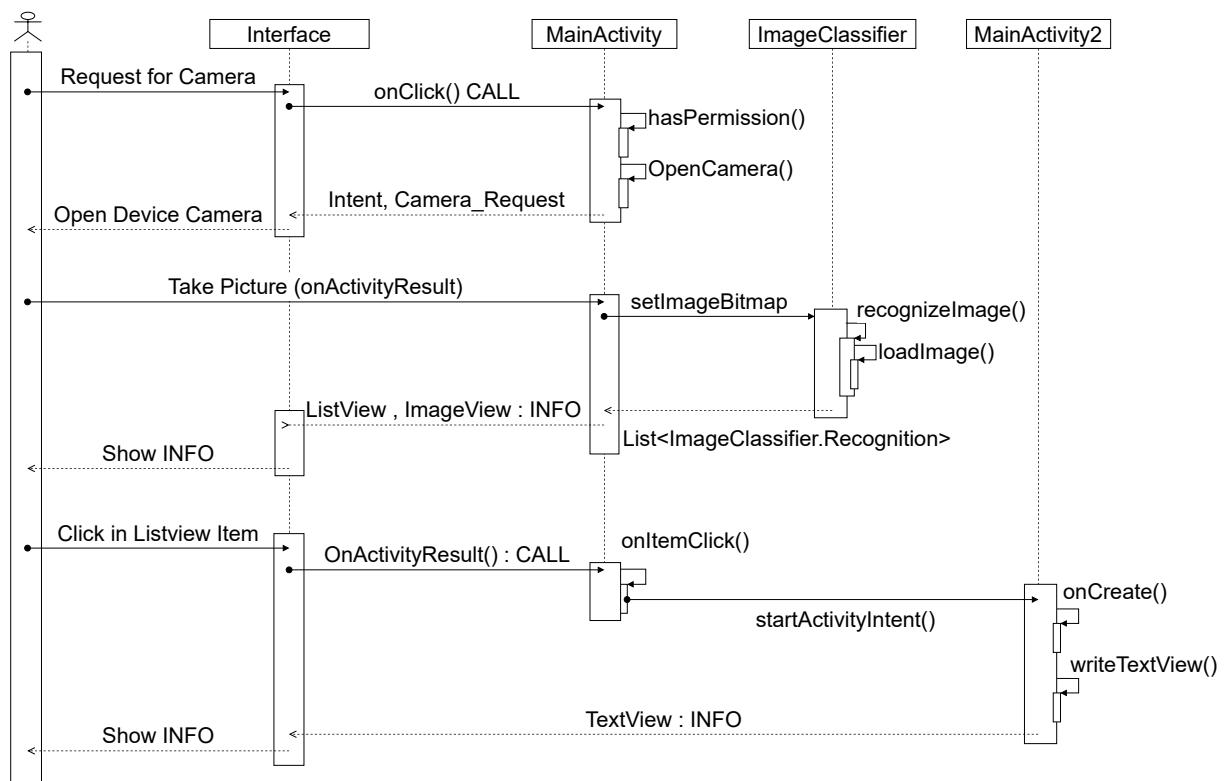
dos dados necessários para a definição do modelo de treino que será utilizado, juntamente com as classes ou “labels” presentes no modelo. Em adição, a “MainActivity” recebe os resultados da classificação e disponibiliza os dados ao usuário através da interface.

A classe “ImageClassifier” é instanciada na “MainActivity” onde são informados o modelo e o label a serem utilizados. Dessa forma, quando o usuário faz a interação de tirar foto, a imagem capturada é transformada para o formato de bitmap e enviada para “ImageClassifier” para fazer a tarefa de classificação. Feita a classificação, os nomes dos labels (nome) e os valores de acurácia (confidence) são salvos em uma sublista de “Recognition”.

6.4 Diagrama de Sequência

Com o objetivo de apresentar, de forma visual, as interações presentes durante o uso da aplicação, a Figura 11 apresenta o diagrama de sequência, que abrange desde a interação inicial do usuário, até a última tela de interação. Dessa forma, o diagrama é constituído pelo ator(Usuário) e pelos componentes(Interface, MainActivity, MainActivity2 e ImageClassifier).

Figura 11. Diagrama de Sequência UML



O diagrama de sequência segue a ordem de interação mostrado na Figura 9 (verificar a seção 6.1), sendo que, o usuário clica no botão da tela inicial, a MainActivity capta à

interação e solicita as permissões necessárias para utilizar a câmera do dispositivo, com as permissões, a câmera é acessada e o usuário tira a foto, em seguida, a foto é captada pela MainActivity e transformada em Bitmap para então ser enviada ao componente ImageClassifier, responsável pela classificação de imagem.

No Componente ImageClassifier a imagem em bitmap passa por diversas transformações com o intuito de seguir os padrões do modelo utilizado para a classificação. Após a classificação, é gerada uma lista com instâncias de Recognition (classe interna a ImageClassifier) para cada um dos *labels* (classes presentes no modelo treinado). Sendo que cada instância de Recognition contém uma variável “String” de chave identificadora e uma variável “float” contendo os valores de acurácia recebidos após a classificação.

Após a criação da lista, esta é apresentada ao usuário juntamente com a foto tirada na segunda tela da aplicação. Com isso, o usuário pode clicar nos itens apresentados em formato de lista na tela, sendo então direcionado para a última tela que apresenta informações históricas sobre o azulejo.

6.5 Utilização do Módulo

Tendo em vista o objetivo de criar um módulo que possa ser reutilizável em outros projetos, que venham a utilizar da classificação de imagens, foram reunidas no módulo as funções necessárias para possibilitar a tarefa de classificação de forma clara, onde o usuário informa apenas as informações sobre as classes e o modelo que será utilizado.

Tendo em vista a facilidade de uso do módulo em outros projetos, é necessário importar o módulo e adicionar algumas dependências necessárias, no Android Studio é possível fazer o “import” na opção de arquivo, como exemplificado na Figura 12. Depois de importar o módulo é necessário adicionar ao *build.gradle* do projeto, as dependências do tensorflow lite, Figura 13. Dessa forma, possibilitando criar uma instância da classe ImageClassifier presente no módulo. Por fim, o programador deve salvar o modelo treinado e o arquivo com as classes na pasta assets do projeto do android studio.

Figura 12. Importando módulo em novo projeto.

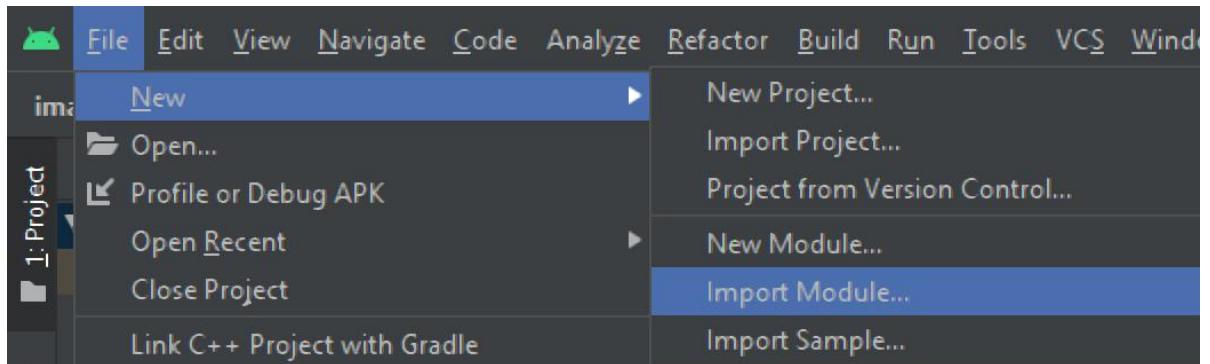
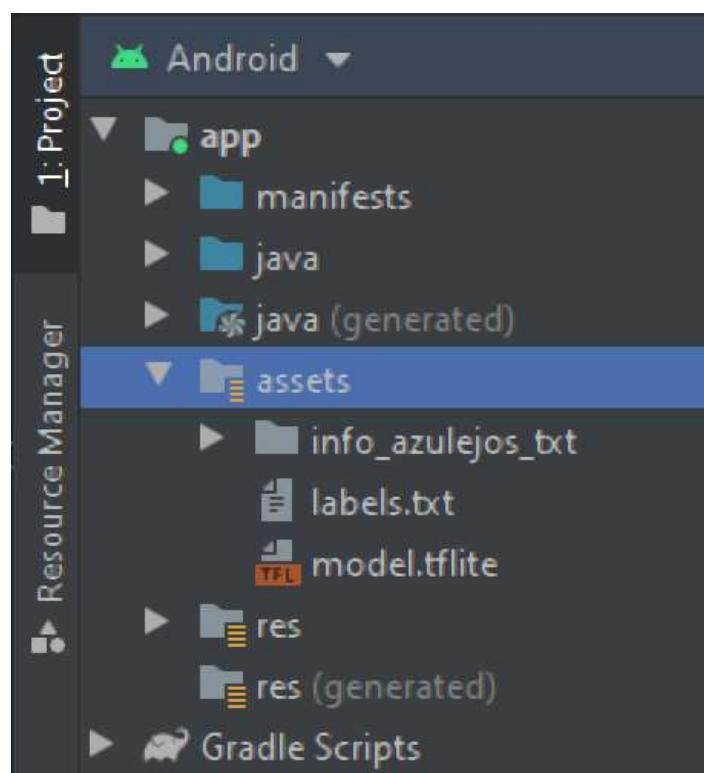


Figura 13. Dependências necessárias no build.gradle do novo projeto.

```
implementation('org.tensorflow:tensorflow-lite:0.0.0-nightly') { changing = true }
implementation('org.tensorflow:tensorflow-lite-gpu:0.0.0-nightly') { changing = true }
implementation('org.tensorflow:tensorflow-lite-support:0.0.0-nightly') { changing = true }
```

Figura 14. Adicionar modelo (model.tflite) e arquivo com classes (labels.txt) no projeto.



7 Conclusão

Este trabalho de conclusão de curso, apresentou a aplicação de uma rede neural convolucional na classificação de azulejos históricos na cidade de Belém do Pará. Para isso, foi criada um módulo de classificação de imagem utilizando as bibliotecas do ambiente Tensorflow, com o objetivo de reutilizar o módulo criado em diferentes projetos que necessitem da tarefa de classificação. Além disso, referente aos azulejos, foi feita a transcrição de uma tabela contendo um inventário da azulejaria de Belém do Pará para o formato digital.

Em referência a aquisição do conhecimento necessário para a execução do trabalho, foi feita uma pesquisa em diversos artigos, visando outras aplicações desenvolvidas com temas semelhantes além de informações mais detalhadas de assuntos que seriam abordados, com o intuito de ter uma visão concreta do que poderia ser desenvolvido.

Em relação a utilização de uma CNN para a tarefa de classificação de imagens, foi utilizada uma arquitetura de dezoito camadas obtida através de uma busca exploratória. A base de dados selecionada para o treinamento, apresenta doze classes de azulejos com um total de 1164 imagens, sendo estas, tiradas de diferentes aparelhos celulares e contando com diferentes ângulos, iluminação, sombra, além de outros elementos. Quanto ao treinamento do modelo, a base de dados foi dividida entre imagens para ao treino e geração do modelo, com 80%, e imagens para o teste do modelo, com 20%. Devido ao número de imagens ser relativamente pequeno, foi utilizada a técnica “Data Augmentation” para a geração de imagens sintéticas. Em adição, devido a indisponibilidade de locomoção decorrente a pandemia de Covid-19, foram inseridas imagem de azulejos tiradas do monitor de computadores, sendo necessário o uso de transformações referentes a interferências de iluminação. Durante o treinamento, o valor da acurácia foi de 98,60% indicando que quase todas as imagens do trino foram classificadas corretamente e o valor de perda de 0,0622%. Durante a fase de testes com a aplicação, os valores de acurácia e perda marcaram os valores de 92,70% e 0,2875% respectivamente, atestando a qualidade do treinamento.

Em relação as contribuições deste trabalho, tendo em vista aplicações desenvolvidas em projetos semelhantes voltados para a questão cultural e turismo, foi decidido pela criação de um módulo de classificação, que reuni as funções e interações necessárias para executar a tarefa de classificação de imagem. Com isso, o módulo foi criado visando apresentar ao programador apenas as ferramentas essenciais para a tarefa.

Em adição, é importante citar a transcrição da tabela presente no inventário da azulejaria de Belém do Pará para o formato digital, contando com 479 imagens de diferentes azulejos, além de diversas informações sobre cada um destes. Dessa forma, a tabela criada

pode ser utilizada tanto em relação as informações contidas, oferecem uma boa fonte de dados para construções de modelos maiores, quanto para a disponibilidade das informações, que podem ser usadas um outros projetos com outras finalidades.

Quanto aos trabalhos futuros, em relação a aplicação possíveis integrações com tecnologia de localização (GPS) e possibilidade de utilização da realidade aumentada. Além da utilização da tabela e suas informações extraídas para criação de modelos de treino maiores. A tabela e informações retiradas, estão disponível online¹.

¹ <<http://labvis.ufpa.br/AzulejAR/>>

Referências

- ALCÂNTARA, D. M. e. S.; BRITO, S. R. S.; SANJAD, T. A. B. C. *Azulejaria em Belém do Pará: Inventário - Arquitetura civil e religiosa - Século XVIII ao XX*. Brasília, DF: Biblioteca Aloísio Magalhães IPHAN, 2016. Citado 4 vezes nas páginas 10, 21, 22 e 23.
- ALFONSO, J. L. M. et al. Geotourism itineraries and augmented reality in the geomorphosites of the arribes del duero natural park (zamora sector, spain). *Geoheritage*, Springer Science and Business Media LLC, v. 13, n. 1, feb 2021. Citado 2 vezes nas páginas 19 e 20.
- ALPAYDIN, E. *Introduction to machine learning*. Cambridge, Mass: MIT Press, 2010. ISBN 9780262012430. Citado na página 11.
- ARRUDA, T. C.; SANJAD, T. A. B. C. Ornamentos de platibanda em edificações de belém entre os séculos XIX e XX: inventário e conservação. *Anais do Museu Paulista: História e Cultura Material*, FapUNIFESP (SciELO), v. 25, n. 3, p. 341–388, sep 2017. Citado na página 10.
- AWANG, K. W.; HASSAN, W. M. W.; ZAHARI, M. S. M. Tourism development: A geographical perspective. *Asian Social Science*, Canadian Center of Science and Education, v. 5, n. 5, apr 2009. Citado na página 10.
- BERGSTRA, J.; BENGIO, Y. Random search for hyper-parameter optimization. *Journal of machine learning research*, v. 13, n. 2, 2012. Citado na página 27.
- BISHOP, C. M. *Pattern Recognition and Machine Learning*. 5. ed. [S.l.]: Springer, 2006. (Information science and statistics). ISBN 978-0-38-731073-2. Citado na página 27.
- BONHAGE, A. et al. A modified mask region-based convolutional neural network approach for the automated detection of archaeological sites on high-resolution light detection and ranging-derived digital elevation models in the north german lowland. *Archaeological Prospection*, Wiley, feb 2021. Citado 2 vezes nas páginas 19 e 20.
- CAUCHI, M.; SCERRI, D. Enriching tourist UX via a location based AR treasure hunt game. In: *2019 IEEE 9th International Conference on Consumer Electronics (ICCE-Berlin)*. [S.l.]: IEEE, 2019. Citado 2 vezes nas páginas 19 e 20.
- CONDE, M. B. Nuevas tecnologías y difusión del turismo cultural: descubriendo a goya con realidad aumentada. *ROTUR. Revista de Ocio y Turismo*, Universidade da Coruna, v. 14, n. 1, p. 81–93, jan 2020. Citado 2 vezes nas páginas 19 e 20.
- COSTA, A. H. M. da. Aplicação de redes neurais convolucionais na identificação de tatuadores. *Departamento de Engenharia Elétrica*, dec 2018. Citado na página 15.
- DEMIR, O. F.; KARAARSLAN, E. Augmented reality application for smart tourism: GökovAR. In: *2018 6th International Istanbul Smart Grids and Cities Congress and Fair (ICSG)*. [S.l.]: IEEE, 2018. Citado 2 vezes nas páginas 19 e 20.

FUKADA, H.; KASAI, K.; OHTSU, S. A field experiment of system to provide tourism information using image recognition type AR technology. In: *Lecture Notes in Electrical Engineering*. [S.l.]: Springer International Publishing, 2014. p. 381–387. Citado 2 vezes nas páginas 19 e 20.

GAMMA, E. et al. *Design Patterns*. [S.l.]: Prentice Hall, 1995. ISBN 0201633612. Citado na página 32.

GODEWITHANA, N. et al. Historical places & monuments identification system. In: *2020 IEEE REGION 10 CONFERENCE (TENCON)*. [S.l.]: IEEE, 2020. Citado 2 vezes nas páginas 19 e 20.

GODOY, R. C. et al. Turismo e realidade aumentada: Desenvolvimento de um aplicativo para a cidade de santiago/rs. *Anais da X edição do Encontro Anual de Tecnologia da Informação - EATI*, v. 9, n. 1, p. 110–117, out. 2019. Disponível em: <<http://anais.eati.info:8080/index.php/2019/article/view/20/17>>. Citado 3 vezes nas páginas 10, 18 e 20.

HAN, S. et al. Extracting representative images of tourist attractions from flickr by combining an improved cluster method and multiple deep learning models. *ISPRS International Journal of Geo-Information*, MDPI AG, v. 9, n. 2, p. 81, jan 2020. Citado 2 vezes nas páginas 19 e 20.

HAYKIN, S. *Neural Networks and Learning Machines*. [S.l.: s.n.], 2008. ISSN 14337851. ISBN 9780131471399. Citado na página 12.

HE, Z.; WU, L.; LI, X. R. When art meets tech: The role of augmented reality in enhancing museum experiences and purchase intentions. *Tourism Management*, Elsevier BV, v. 68, p. 127–139, oct 2018. Citado 2 vezes nas páginas 19 e 20.

HOANG, V.-D. et al. Deep feature extraction for panoramic image stitching. In: *Intelligent Information and Database Systems*. [S.l.]: Springer International Publishing, 2020. p. 141–151. Citado na página 20.

HUANG, T.-L. Restorative experiences and online tourists' willingness to pay a price premium in an augmented reality environment. *Journal of Retailing and Consumer Services*, Elsevier BV, v. 58, p. 102256, jan 2021. Citado 2 vezes nas páginas 19 e 20.

INDOLIA, S. et al. Conceptual understanding of convolutional neural network- a deep learning approach. *Procedia Computer Science*, Elsevier BV, v. 132, p. 679–688, 2018. Citado na página 11.

JURASZEK, G. *Image Recognition of Products Using Bag of Visual Words and Convolutional Neural Networks*. [S.l.]: Unpublished, 2014. Citado na página 18.

KOUNAVIS, C. D.; KASIMATI, A. E.; ZAMANI, E. D. Enhancing the tourism experience through mobile augmented reality: Challenges and prospects. *International Journal of Engineering Business Management*, SAGE Publications, v. 4, p. 10, jan 2012. Citado na página 10.

KRIZHEVSKY, A.; SUTSKEVER, I.; HINTON, G. E. Imagenet classification with deep convolutional neural networks. In: *Advances in neural information processing systems*. [S.l.: s.n.], 2012. p. 1097–1105. Citado na página 27.

- KYSELA, J.; ŠTORKOVÁ, P. Using augmented reality as a medium for teaching history and tourism. *Procedia - Social and Behavioral Sciences*, Elsevier BV, v. 174, p. 926–931, feb 2015. Citado 2 vezes nas páginas 19 e 20.
- LAMBERS, K.; VAART, W. V. van der; BOURGEOIS, Q. Integrating remote sensing, machine learning, and citizen science in dutch archaeological prospection. *Remote Sensing*, MDPI AG, v. 11, n. 7, p. 794, apr 2019. Citado 2 vezes nas páginas 19 e 20.
- LECUN, Y.; BENGIO, Y.; HINTON, G. Deep learning. *Nature*, Nature Publishing Group, v. 521, n. 7553, p. 436–444, 2015. Citado 2 vezes nas páginas 14 e 15.
- LECUN, Y. et al. Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, Ieee, v. 86, n. 11, p. 2278–2324, 1998. Citado 2 vezes nas páginas 6 e 16.
- LEI, Q. et al. Research on image recognition method of ethnic costume based on VGG. In: *Machine Learning for Cyber Security*. [S.l.]: Springer International Publishing, 2020. p. 312–325. Citado 2 vezes nas páginas 19 e 20.
- LIU, W. et al. A survey of deep neural network architectures and their applications. *Neurocomputing*, Elsevier BV, v. 234, p. 11–26, apr 2017. Citado na página 14.
- LLERENA, J.; ANDINA, M.; GRIJALVA, J. Mobile application to promote the malecón 2000 tourism using augmented reality and geolocation. In: *2018 International Conference on Information Systems and Computer Science (INCISCOS)*. [S.l.]: IEEE, 2018. Citado 2 vezes nas páginas 19 e 20.
- LUCENA, J. O. B. A. de. Captura de movimento facial em tempo real para realidade aumentada. *Dissertação (Mestrado em Ciência da Computação) – Universidade Federal de Pernambuco, Recife*, aug 2019. Citado na página 11.
- MARTINS, M.; MALTA, C.; COSTA, V. Viseu mobile: A location based augmented reality tour guide for mobile devices. *Dos Algarves: a Multidisciplinary e-journal*, School of Management, Hospitality and Tourism, University of the Algarve, v. 26, n. 1, p. 8–26, nov 2015. Citado 2 vezes nas páginas 11 e 19.
- MARTO, A. et al. Development and evaluation of an outdoor multisensory AR system for cultural heritage. *IEEE Access*, Institute of Electrical and Electronics Engineers (IEEE), v. 9, p. 16419–16434, 2021. Citado 2 vezes nas páginas 19 e 20.
- MEDINA, M. *Algoritmos e programacao: teoria e pratica*. Sao Paulo: Novatec, 2006. ISBN 9788575220733. Citado na página 32.
- MOHRI, M. *Foundations of machine learning*. Cambridge, Massachusetts: The MIT Press, 2018. ISBN 0262351366. Citado na página 14.
- NEVES, L. A. P.; NETO, H. V.; GONZAGA, A. *Avanços em Visão Computacional*. [S.l.]: Omnipax Editora Ltda, 2012. Citado na página 10.
- O'SHEA, K.; NASH, R. An introduction to convolutional neural networks. nov. 2015. Citado 4 vezes nas páginas 11, 14, 15 e 16.
- PANKANTI, S. et al. Practical computer vision: Example techniques and challenges. *IBM Journal of Research and Development*, IBM, v. 55, n. 5, p. 3:1–3:12, sep 2011. Citado na página 11.

- POUX, F. et al. Initial user-centered design of a virtual reality heritage system: Applications for digital tourism. *Remote Sensing*, MDPI AG, v. 12, n. 16, p. 2583, aug 2020. Citado 2 vezes nas páginas 19 e 20.
- REIS, C. F.; ALBUQUERQUE, M. P. de; CASTRO, S. B. de. Introdução ao reconhecimento de padrões utilizando redes neurais. *Centro Brasileiro de Pesquisas Físicas – CBPF-NT-002/01*, 2001. Citado na página 11.
- ROCHA, P. M. A. A exploração da realidade aumentada pelo jornalismo: a exposição da informação dos média num espaço aumentado. *CECS - Centro de Estudos de Comunicação e Sociedade Universidade do MinhoBraga . Portugal. Literacia, Media e Cidadania - Livro de Atas do 4.º Congresso*, p. 475–491, nov 2017. Citado na página 18.
- ROJAS, R. *Neural networks : a systematic introduction*. Berlin New York: Springer-Verlag, 1996. ISBN 9783642610684. Citado na página 14.
- SANTOS, C. et al. Recognizing and exploring azulejos on historic buildings facades by combining computer vision and geolocation in mobile augmented reality applications. *Journal of Mobile Multimedia*, 2017. Disponível em: <<https://dl.acm.org/doi/abs/10.5555/3177197.3177201>>. Citado 4 vezes nas páginas 12, 23, 24 e 25.
- SILVA, L. et al. Redes neurais convolucionais aplicadas na detecção de pneumonia através de imagens de raio-x. In: *Anais do 14. Congresso Brasileiro de Inteligência Computacional*. [S.I.]: ABRICOM, 2020. Citado 2 vezes nas páginas 11 e 18.
- TRIER, Ø. D.; REKSTEN, J. H.; LØSETH, K. Automated mapping of cultural heritage in norway from airborne lidar data using faster r-CNN. *International Journal of Applied Earth Observation and Geoinformation*, Elsevier BV, v. 95, p. 102241, mar 2021. Citado 2 vezes nas páginas 19 e 20.
- WILLIAMS, A. A. L. S. *Tourism Geography: Critical Understandings of Place, Space and Experience*. [S.I.]: Taylor & Francis Ltd., 2014. ISBN 9781135010171. Citado na página 10.
- ZENDEL, O. et al. Analyzing computer vision data - the good, the bad and the ugly. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. [S.I.: s.n.], 2017. p. 1980–1990. Citado na página 11.
- ZHOU, X. Understanding the convolutional neural networks with gradient descent and backpropagation. *Journal of Physics: Conference Series*, IOP Publishing, v. 1004, p. 012028, apr 2018. Citado na página 17.