

# Extração Automática de Metadados de Imagens de Acervos Digitais de Museus Brasileiros

Vagner Inácio de Oliveira , Paula D. Paro Costa , Dalton Martins  
{vagner.inol@gmail.com,paulad@unicamp.br,daltonmartins@unb.br}

Departamento de Engenharia de Computação e Automação (DCA)  
Faculdade de Engenharia Elétrica e de Computação (FEEC)  
Universidade Estadual de Campinas (Unicamp)  
Campinas, SP, Brasil

**Abstract** – A anotação de metadados de acervos digitais é realizada, tipicamente, por diferentes profissionais especializados, configurando uma atividade complexa, trabalhosa, que demanda grande quantidade de tempo, frequentemente sujeita a falhas humanas, altos custos e problemas na recuperação das informações de acordo com o desejado. Avanços recentes em inteligência artificial, particularmente técnicas de *Deep Learning*, têm mostrado seu potencial na realização de reconhecimentos visuais e na interpretação de objetos em imagens. Nesse contexto, o presente trabalho apresenta o EMA, um conjunto de dados de imagens oriundas do patrimônio cultural brasileiro com mais de 11.000 imagens rotuladas de objetos pertencentes a dezessete museus brasileiros, disponibilizadas pelo Projeto Tainacan. O conjunto de dados EMA é uma contribuição para o desenvolvimento de ferramentas de anotação de metadados automatizadas. Este projeto também apresenta resultados parciais da rede neural residual ResNet50 como *baseline* para o conjunto de dados, resultando em uma taxa de reconhecimento superior a 86%.

**Keywords** – Patrimônios Culturais Digitais, Tesouro, Anotação Automática, Deep Learning, Computer Vision

## 1. Introdução

As coleções digitais são uma maneira eficaz de possibilitar ao público a exploração do patrimônio cultural dos museus. Elas são particularmente relevantes em um país como o Brasil, onde os museus que preservam a história do país estão a milhares de quilômetros de distância, tornando-os inacessíveis à maioria das pessoas e difíceis de serem estudados por historiadores e pesquisadores em geral. Adiciona-se a este fator o risco de desastres, vide os ocorridos em menos de uma década com três museus brasileiros: Museu da Língua Portuguesa em São Paulo em 2015, o Museu Histórico Nacional no Rio de Janeiro em 2018 e, mais recentemente, o Museu de História Natural em Minas Gerais em 2020.

Apesar de todas as dificuldades enfrentadas pelos museus brasileiros, o país possui uma quantidade expressiva de acervos digitalizados. O Instituto Brasileiro de Museus (IBRAM) dá acesso pela internet a mais de 15.000 itens, de dezessete museus, em conjunto com seus respectivos metadados anotados com contexto histórico. A principal tecnologia da informação por trás disso é o Tainacan [3], uma plataforma de código aberto para criação de acervos digitais no WordPress, que também permite acesso programável ao banco de dados [1].

A anotação de metadados completa e confiável é fundamental para agregar significado às imagens do acervo digital de um museu. A imagem de um garfo, por exemplo, torna-se uma imagem irrelevante de um objeto se

não for indicado que foi utilizado por algum personagem histórico durante um jantar onde foram tomadas grandes decisões ou que seu material representa todo um período histórico. Essa anotação de metadados é normalmente conduzida por vários profissionais especializados e é uma atividade complexa, trabalhosa e demorada, frequentemente levando a altos custos, falhas humanas e mal-entendidos. Com isso, inúmeros acervos digitalizados no Brasil e no mundo sofrem com a falta de informações de metadados, tornando os bens culturais pouco atrativos e seu potencial completo inexplorado.

Para enfrentar o problema, este trabalho propõe o uso de algoritmos de aprendizado de máquina, especificamente, modelos de *computer vision*, como ferramentas de auxílio para que profissionais especializados conduzam processos de anotação de metadados mais eficientes, confiáveis e potencialmente menos dispendiosos e descreve a construção de um conjunto de dados de imagens como um passo necessário para o desenvolvimento de ferramentas de anotação de metadados baseadas em IA para bens de patrimônios culturais.

## 2. Proposta

O principal objetivo deste projeto será a criação de uma base de dados confiável de imagens rotuladas para uma posterior aplicação no desenvolvimento de um modelo de extração automática de metadados. A validação do conjunto de dados será efetuada através de um modelo que

servirá como *baseline*.

### 3. Metodologia

O primeiro passo em nossa metodologia envolveu um estudo do acervo digital gerenciado pelo IBRAM integrado pelo Tainacan [3]. Coletamos os metadados de todos os objetos da coleção, que contém 15.651 objetos de dezessete museus (“JSON Metadata” na Figura 1).

Cada objeto do acervo é categorizado de acordo com seu tesouro. Um tesouro é definido como um conjunto de conceitos, denominados termos ou descritores, determinados de acordo com sua função ou estrutura, ordenados de forma clara e inequívoca, com base no estabelecimento de relações entre eles [2].

Como primeira abordagem do problema, focamos no tesouro mais frequente na coleção, “interior”, correspondendo a 18,6% do total de itens. No contexto de patrimônio cultural, o termo refere-se a objetos da vida cotidiana usados no interior das casas, tal como um ferro a carvão usado para passar roupas quando não havia eletricidade. Também realizamos uma entrevista com uma museóloga do IBRAM que confirmou que muitos museus no Brasil se dedicam a mostrar como as pessoas viviam no passado, mostrando, por exemplo, como brancos e negros viviam na época da escravidão. Ela também enfatizou a relevância de desenvolver ferramentas automáticas ou semiautomáticas para ajudar os museólogos a gerar metadados para itens digitalizados.

Como segundo passo, analisamos qual campo de metadados poderia ser utilizado para rotular suas imagens correspondentes. Identificamos que os campos de metadados “título”, “denominação”, “tipo de material” e “técnica” são os que fornecem uma descrição geral de cada item. No entanto, verificamos que os campos “tipo de material” e “técnica” nem sempre eram preenchidos e que o campo “título” às vezes substituíra uma descrição precisa por um alias que não descreve o objeto adequadamente. Por esse motivo, adotamos o campo “denominação” como o campo de destino para extrair os rótulos do nosso conjunto de dados.

Mais uma vez, nos deparamos com uma vasta quantidade de termos usados para descrever os objetos “interiores” da coleção, e decidimos analisar as palavras utilizadas com maior frequência para descrever os objetos. Como resultado dessa análise, decidimos manter apenas as 31 palavras mais frequentes como rótulos das imagens, se tornando estes, o nome das pastas no banco de dados de imagens, pois também facilita a sua utilização no modelo. Alguns exemplos de objetos “interiores” obtidos da coleção e seus respectivos rótulos podem ser vistos na

Figura 2.



Figura 2. Exemplos de objetos pertencentes ao tesouro “interior”. Seus rótulos são castiçal (*candlestick*), panela (*pan*) e lampião (*kerosene lamp*).

### 4. Resultados

O principal resultado do presente trabalho é o conjunto de dados de imagens EMA, com 11.996 imagens, correspondentes a 2.922 objetos de dezessete museus brasileiros, rotulados de acordo com 31 classes. Como prova de conceito de uso do conjunto de dados EMA para treinar um modelo de *Deep Learning* (DL) para reconhecimento de objetos de “interiores” de patrimônios culturais, construímos um classificador de imagens que conta com a rede pré-treinada ResNet50. Adotando o método de aprendizagem por transferência, treinamos a camada final usando as imagens originais sem aumento de dados ou quaisquer transformações. Usamos 80% das imagens para treinar o modelo e as imagens restantes foram usadas para validação e testes. O modelo foi aplicado com fastai, uma biblioteca de DL de código aberto construída em cima do PyTorch.

A precisão de treinamento e validação ao final de 6 épocas foi de 86,7%. As classificações mais confusas estão resumidas na Tabela 3 e mostram as limitações de nossa metodologia. Por exemplo, os resultados dos quatro rótulos para identificar talheres: garfo, faca de mesa, colher e também talheres. Esses quatro rótulos resultaram em muitos erros de classificação, já que o rótulo dos talheres engloba garfo, faca e colher. Notamos também, por exemplo, a confusão entre as classes luminária e arandela. Uma luminária pode ter partes de uma arandela, por isso não é fácil resolver esse tipo de classificação.

### 5. Conclusões

A anotação de metadados em coleções digitais é uma tarefa desafiadora. Os problemas típicos incluem falta de informação e classificações erradas, principalmente devido a diferenças significativas entre objetos modernos e seus equivalentes no passado. Esses problemas podem causar problemas de recuperação de dados ou associar um item ao contexto errado, dificultando o acesso ao conhecimento que o objeto pode oferecer.

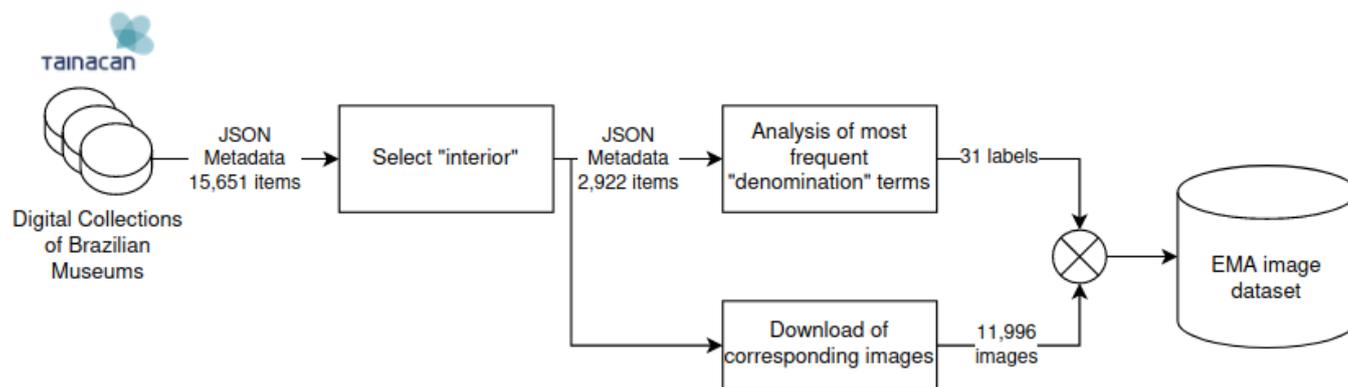


Figura 1. Passos para a construção do conjunto de dados EMA.

Actual	Predicted	occurrences
spoon	cutlery	58
table knife	cutlery	44
cutlery	spoon	34
cutlery	table knife	32
sideboard	curtain	31
luminaire	sconce	26
sconce	luminaire	17
cutlery	fork	16
curtain	sideboard	15
fork	cutlery	11
bed	jug	7
fork	spoon	5
dish	spoon	4
cup of tea	spoon	3
mirror	chest	3
saucer	cup of tea	3
sideboard	table	3
chest	mirror	2
cup	spoon	2
cup of tea	saucer	2
luminaire	table	2
table	luminaire	2
table	sideboard	2
chest of drawers	chest	1
chest of drawers	fork	1
glass	spoon	1
lamp	jug	1
luminaire	table knife	1
table knife	spoon	1

Figura 3. Confusões com maior frequência

Neste projeto, apresentamos nossos primeiros passos para o desenvolvimento de ferramentas de anotação de metadados baseadas em IA para ajudar os museólogos a melhorar a qualidade geral da anotação de coleções digitais. Em particular, apresentamos o EMA, um conjunto de dados de imagens rotuladas com mais de 11.000 imagens de objetos históricos encontrados em dezessete museus brasileiros. O código implementado para executar todas as etapas de processamento e classificação descritas

neste trabalho e as instruções para solicitação do conjunto de dados estão disponíveis no repositório do projeto [4].

Também apresentamos resultados de um modelo *baseline* para este conjunto de dados por meio de um treinamento do modelo ResNet50. Nosso modelo conseguiu obter 86,7% de precisão no reconhecimento de objetos, mostrando a consistência do conjunto de dados e o potencial dessa abordagem.

Trabalhos futuros incluem explorar o desempenho de outras arquiteturas de DL e aumentar o conjunto de dados com outras coleções de patrimônios culturais para uma generalização do modelo. Também planejamos desenvolver um aplicativo que indicará rótulos durante os processos de anotação.

### Agradecimentos

Agradecemos a Amanda Oliveira, do IBRAM, pelas valiosas informações que forneceu sobre sua experiência nos museus brasileiros. Agradecemos também à comunidade Tainacan pelo apoio.

### Referências

- [1] Governo Federal. Instituto brasileiro de museus - ibram, 2022.
- [2] Helena Dodd Ferrez. Tesouro de objetos do patrimônio cultural nos museus brasileiros. *Rio de Janeiro: Fazer Arte. Gerência de Museus da Secretaria Municipal de Cultura*, 2016.
- [3] University of Brasília. Tainacan, 2021.
- [4] P. D. P. Costa e D. L. Martins V. De Oliveira. "ema's project repository", 2022.