Classificação automatizada de obras de arte utilizando o modelo clip: uma análise comparativa entre artistas brasileiros e europeus do século XVI ao XIX

Vitor Amadeu Souza[1]; 0009-00-02-1857-6799

1 – UniFOA, Centro Universitário de Volta Redonda, Volta Redonda, RJ.
vitor.amadeu@foa.org.br

Resumo: Este estudo investiga a aplicabilidade do modelo CLIP (Contrastive Language-Image Pre-training) na classificação automatizada de obras de arte, focando na distinção entre estilos de artistas brasileiros e europeus dos séculos XVI ao XIX. O modelo CLIP foi aplicado na classificação de três obras emblemáticas: Mona Lisa (Leonardo da Vinci), A Primeira Missa no Brasil (Victor Meirelles) e Fala do Trono (Pedro Américo). A metodologia empregou técnicas de zero-shot learning, permitindo classificações sem treinamento específico prévio para o domínio artístico. Os resultados demonstraram alta precisão na identificação de Leonardo da Vinci (99,99% de confiança), moderada precisão para Victor Meirelles (53,57%) e elevada precisão para Pedro Américo (86,54%). Os achados sugerem que o modelo CLIP apresenta eficácia variável dependendo das características estilísticas distintivas de cada artista, com maior precisão para artistas com estilos mais consolidados e reconhecíveis globalmente. Este trabalho contribui para a crescente área de aplicação de inteligência artificial em estudos de arte e patrimônio cultural, oferecendo insights sobre as limitações e potencialidades dos modelos multimodais na análise automatizada de obras artísticas.

Palavras-chave: CLIP. Classificação de imagens. Arte. Aprendizado de máquina. Visão computacional. Zero-shot learning.

INTRODUÇÃO

A intersecção entre inteligência artificial e estudos artísticos tem emergido como um campo de pesquisa promissor, oferecendo novas perspectivas para a análise, classificação e compreensão de obras de arte. O desenvolvimento de modelos multimodais, que processam simultaneamente informações visuais e textuais, revolucionou as possibilidades de aplicação de técnicas computacionais no domínio cultural e artístico. O modelo CLIP (Contrastive Language-Image Pre-training), desenvolvido pela OpenAI, representa um marco significativo nesta evolução tecnológica. Diferentemente dos modelos tradicionais de classificação de imagens, que requerem treinamento específico para cada categoria, o CLIP utiliza aprendizado contrastivo para estabelecer correspondências entre representações visuais e textuais, permitindo classificações zero-shot em domínios não vistos durante o treinamento (Radford *et al.*, 2021). Esta característica torna o modelo particularmente atrativo para aplicações em arte, onde a diversidade de estilos, períodos e técnicas artísticas apresenta desafios únicos para sistemas de classificação automatizada.

A aplicação de técnicas de aprendizado de máquina na análise de arte não é recente. Johnson *et al*. (2008) exploraram o uso de redes neurais para classificação de estilos artísticos, estabelecendo precedentes importantes para pesquisas subsequentes. Saleh e Elgammal (2015) desenvolveram sistemas capazes de identificar períodos artísticos com base em características visuais extraídas automaticamente, enquanto Cetinic *et al*. (2018) investigaram a eficácia de diferentes arquiteturas de deep learning na classificação de movimentos artísticos. Mais recentemente, Elgammal *et al.* (2017) introduziram redes adversárias criativas (CANs) que não apenas classificam, mas também geram arte original baseada no aprendizado de estilos existentes.

No contexto brasileiro, a arte dos séculos XVIII e XIX apresenta características particulares que refletem tanto influências europeias quanto elementos culturais autóctones. Victor Meirelles (1832-1903) e Pedro Américo (1843-1905) destacam-se como figuras centrais do romantismo brasileiro, desenvolvendo obras que combinam técnicas acadêmicas europeias com temáticas nacionais. A comparação destes artistas com mestres europeus como Leonardo da Vinci (1452-1519) oferece uma oportunidade única para avaliar a capacidade

do modelo CLIP em distinguir características estilísticas tanto temporais quanto geográficas. Stork (2009) argumenta que a análise computacional de pinturas requer consideração cuidadosa de fatores como digitalização, iluminação e preservação, aspectos particularmente relevantes ao comparar obras de diferentes períodos históricos.

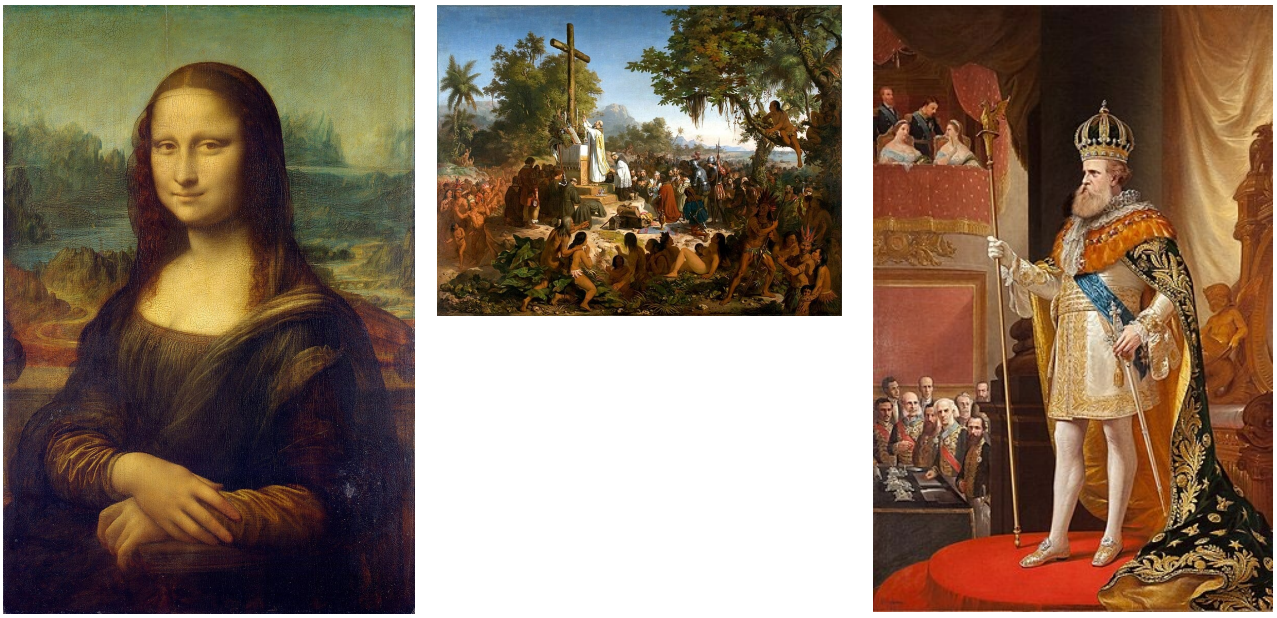
O presente estudo objetiva avaliar a eficácia do modelo CLIP na classificação automatizada de obras de arte, investigando sua capacidade de distinguir entre estilos de diferentes artistas e períodos históricos. Especificamente, busca-se implementar um sistema de classificação zero-shot utilizando o modelo CLIP, avaliar a precisão do modelo na identificação de obras de Leonardo da Vinci, Victor Meirelles e Pedro Américo, analisar os fatores que influenciam a confiança das classificações, e discutir as implicações dos resultados para futuras aplicações em estudos de arte digital.

MÉTODOS

O experimento utilizou o modelo CLIP na versão "openai/clip-vit-base-patch32", que combina um codificador visual baseado em Vision Transformer (ViT) com um codificador textual. O modelo ViT-Base emprega patches de 32x32 pixels, processando imagens de resolução 224x224 através de 12 camadas de transformers com dimensão de embedding de 768. A escolha desta arquitetura baseou-se em estudos que demonstram sua eficácia em tarefas de classificação zero-shot, particularmente em domínios culturais (Radford *et al*., 2021). O processamento das imagens seguiu o protocolo padrão do CLIP, incluindo redimensionamento, normalização e tokenização conforme especificações do modelo pré-treinado.

Foram selecionadas três obras representativas de diferentes períodos e estilos artísticos: Mona Lisa de Leonardo da Vinci, exemplar do Renascimento italiano; A Primeira Missa no Brasil de Victor Meirelles, representativa do romantismo brasileiro; e Fala do Trono de Pedro Américo, característica do academicismo brasileiro do século XIX. A seleção priorizou obras em domínio público disponíveis em repositórios digitais de alta qualidade, especificamente a Wikimedia Commons, garantindo acesso livre e resolução adequada para processamento pelo modelo. As imagens foram obtidas em formato RGB com resolução mínima de 800x600 pixels. A Figura 1 apresenta as imagens usadas como referência.

Figura 1 - Monalisa (Leonardo da Vinci), A Primeira Missa no Brasil (Victor Meirelles) e Fala do Trono (Pedro Américo)



Fonte: Wikimedia.

O algoritmo de classificação implementou as seguintes etapas: extração de características visuais utilizando o codificador visual do CLIP para gerar embeddings das imagens de entrada; codificação textual processando os nomes dos artistas candidatos através do codificador textual; normalização aplicando normalização L2 aos embeddings visuais e textuais para garantir comparabilidade; cálculo de similaridade computando o produto escalar entre embeddings normalizados, escalado por fator de 100 para amplificação das diferenças; e classificação probabilística aplicando função softmax para conversão de pontuações de similaridade em distribuição probabilística.

A implementação utilizou Python com as bibliotecas Transformers da Hugging Face para acesso ao modelo CLIP, PyTorch para operações tensoriais, PIL para processamento de imagens, e Requests para download de imagens. O código foi executado em ambiente CPU, com processamento sequencial das três obras selecionadas. A avaliação utilizou duas métricas principais: precisão da classificação, determinada pela correspondência entre o artista predito e o verdadeiro autor da obra; e confiança da classificação, expressa como probabilidade softmax da classe predita. Adicionalmente, foram analisadas as distribuições

de probabilidade para todas as classes candidatas, fornecendo insights sobre a discriminabilidade entre diferentes artistas.

O código-fonte está disponível para download através do link: https://github.com/vitor-souza-ime/arte.

RESULTADOS E DISCUSSÃO

Os resultados obtidos revelam variações significativas na capacidade do modelo CLIP de classificar obras de diferentes artistas. A Mona Lisa foi classificada corretamente como obra de Leonardo da Vinci com confiança de 99,99%. A Primeira Missa no Brasil foi identificada como obra de Victor Meirelles com confiança moderada de 53,57%. A Fala do Trono foi classificada corretamente como obra de Pedro Américo com confiança elevada de 86,54%. Todas as classificações foram precisas, demonstrando a capacidade do modelo CLIP em identificar corretamente os autores das obras analisadas.

O resultado obtido para a Mona Lisa corrobora estudos anteriores que demonstram a eficácia de modelos de visão computacional na identificação de obras renascentistas. Este resultado pode ser atribuído a múltiplos fatores: a ampla representação de obras de Leonardo da Vinci nos conjuntos de treinamento de modelos de larga escala; as características técnicas distintivas do sfumato leonardesco e a iconicidade global da obra, que resulta em maior presença em bases de dados visuais na internet (Gatys *et al.*, 2016). A técnica do sfumato, característica fundamental do estilo de Leonardo da Vinci, envolve transições graduais entre cores e tons, criando efeitos de profundidade e atmosfera únicos. Estas características visuais distintivas podem facilitar a identificação automatizada, uma vez que representam padrões consistentes reconhecíveis por algoritmos de aprendizado profundo.

A confiança relativamente baixa na classificação da obra de Victor Meirelles sugere limitações do modelo CLIP em distinguir artistas brasileiros do século XIX. Este resultado alinha-se com observações sobre dificuldades de modelos pré-treinados em classificar arte de regiões sub-representadas nos conjuntos de treinamento (Cetinic *et al.*, 2018). Victor Meirelles desenvolveu um estilo que combina influências neoclássicas europeias com temáticas nacionais brasileiras. Esta hibridização estilística pode criar ambiguidades para

modelos treinados predominantemente em arte europeia, resultando em classificações menos confiantes. Adicionalmente, a menor presença de obras de Meirelles em repositórios digitais internacionais pode limitar a capacidade do modelo de aprender características distintivas deste artista.

A classificação de Pedro Américo apresentou confiança substancial, sugerindo que o modelo conseguiu capturar características estilísticas distintivas do artista. Pedro Américo, formado na École des Beaux-Arts de Paris, desenvolveu um estilo acadêmico rigoroso que pode apresentar maior similaridade com padrões europeus presentes nos dados de treinamento do CLIP. A obra "Fala do Trono" exemplifica o academicismo brasileiro, caracterizado por composições complexas, uso dramático da luz e precisão técnica no desenho. Estas características, alinhadas com tradições artísticas europeias amplamente representadas em conjuntos de dados de treinamento, podem explicar a maior confiança na classificação.

Os resultados evidenciam tanto potencialidades quanto limitações dos modelos multimodais atuais na análise de arte. A alta precisão para artistas canônicos como Leonardo da Vinci contrasta com desafios na identificação de artistas de tradições menos representadas globalmente. Este viés de representação reflete limitações mais amplas dos modelos de aprendizado de máquina, que tendem a reproduzir desigualdades presentes nos dados de treinamento (Bolukbasi *et al*., 2016). Para aplicações práticas em museus e instituições culturais, estes resultados sugerem a necessidade de desenvolvimento de modelos especializados ou fine-tuning com conjuntos de dados mais diversificados e representativos da arte global.

CONCLUSÕES

Este estudo demonstrou a aplicabilidade do modelo CLIP na classificação automatizada de obras de arte, revelando variações significativas na eficácia dependendo das características dos artistas e obras analisadas. Os resultados obtidos confirmam que modelos multimodais pré-treinados podem ser eficazes para identificação artística, mas evidenciam limitações importantes relacionadas à representatividade dos dados de treinamento. A alta confiança na classificação de Leonardo da Vinci contrasta com desafios na identificação de artistas

brasileiros como Victor Meirelles, sugerindo vieses de representação que refletem desigualdades históricas na digitalização e disponibilização de patrimônio cultural.

Estes achados têm implicações importantes para a aplicação de inteligência artificial em contextos culturais diversos, destacando a necessidade de desenvolvimento de abordagens mais inclusivas e representativas. Para pesquisas futuras, recomenda-se expansão do conjunto de dados para incluir maior diversidade de artistas e períodos, investigação de técnicas de fine-tuning para melhorar performance em contextos culturais específicos, desenvolvimento de métricas de avaliação que considerem aspectos qualitativos da análise artística, e exploração de abordagens híbridas que combinem análise automatizada com expertise humana.

O potencial transformador da inteligência artificial nos estudos de arte é evidente, mas sua implementação responsável requer consciência crítica sobre limitações, vieses e implicações culturais. Este trabalho contribui para este diálogo essencial, oferecendo evidências empíricas que podem orientar desenvolvimentos futuros na interseção entre tecnologia e patrimônio cultural, demonstrando que enquanto o modelo CLIP apresenta capacidades promissoras para classificação automatizada de arte, sua aplicação requer consideração cuidadosa de questões de representatividade e contextualização cultural.

REFERÊNCIAS

BOLUKBASI, T. et al. Man is to computer programmer as woman is to homemaker? debiasing word embeddings. Advances in Neural Information Processing Systems, v. 29, p. 4349-4357, 2016.

CETINIC, E. et al. Fine-tuning convolutional neural networks for fine art classification. Expert Systems with Applications, v. 114, p. 107-118, 2018.

ELGAMMAL, A. et al. CAN: Creative adversarial networks, generating "art" by learning about styles and deviating from style norms. 8th International Conference on Computational Creativity, 2017.

GATYS, L. A.; ECKER, A. S.; BETHGE, M. Image style transfer using convolutional neural networks. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, p. 2414-2423, 2016.

JOHNSON, C. R. et al. Image processing for artist identification. IEEE Signal Processing Magazine, v. 25, n. 4, p. 37-48, 2008.

RADFORD, A. et al. Learning transferable visual models from natural language supervision. International Conference on Machine Learning, p. 8748-8763, 2021.

SALEH, B.; ELGAMMAL, A. Large-scale classification of fine-art paintings: Learning the right metric on the right feature. International Journal for Digital Art History, n. 2, p. 71-93, 2015.

STORK, D. G. Computer vision and computer graphics analysis of paintings and drawings: An introduction to the literature. International Conference on Computer Analysis of Images and Patterns, p. 9-24, 2009.