

AVALIANDO O IMPACTO DE APLICAÇÃO DE TÉCNICAS DE LIMIAZIZAÇÃO EM RECONHECIMENTO ÓTICO DE MÚSICA

EVALUATING DE IMPACT IN APPLYING THRESHOLDING TECHNIQUES ON
OMR

Ciências Exatas e da Terra, Engenharias • 31/08/2026

REGISTRO DOI: [10.70773/revistatopicos/788035564](https://doi.org/10.70773/revistatopicos/788035564)

Gustavo Henrique Romão

Hygor Santiago

Jesuliana Nascimento Ulysses Guimarães

Érica da Costa Reis Carvalho

RESUMO

A preservação do patrimônio musical escrito esbarra na progressiva degradação física de partituras históricas, comprometendo sua conservação, acesso e transmissão às futuras gerações. Optical Music Recognition (OMR) é uma tecnologia baseada em inteligência artificial e visão computacional capaz de identificar, interpretar e converter partituras impressas ou manuscritas em formatos digitais editáveis ou reproduzíveis. Embora seja fundamental para a digitalização desses acervos, o ruído decorrente de manchas, dobras e do envelhecimento do papel prejudica o desempenho de modelos de detecção baseados em deep learning. Este artigo investiga a aplicação de técnicas clássicas de limiarização (Otsu e Sauvola) como etapa de pré-processamento para melhorar a robustez do detector YOLOv11 na identificação de elementos musicais em partituras manuscritas degradadas. Através de um pipeline experimental comparativo, demonstrou-se que o método adaptativo de Sauvola, em particular, eleva significativamente a performance do modelo. Os resultados mostram um aumento de 14,3 pontos percentuais no mAP@0.5 (de 64,2% para 78,5%) em relação à imagem original não processada.

Palavras-chave: Reconhecimento ótico de música; Preservação Digital; Partituras Históricas; Limiarização; Binarização; YOLOv11; Visão Computacional; Seleção de atributos.

ABSTRACT

The preservation of written musical heritage faces the challenge of physical degradation of historical scores. Optical Music Recognition (OMR) is crucial for their digitization, but noise from stains, folds, and paper aging compromises the performance of deep learning-based detection models. This paper investigates the application of classical thresholding techniques (Otsu and Sauvola) as a preprocessing step

to improve the robustness of the YOLOv11 detector in identifying musical elements in degraded handwritten scores. Through a comparative experimental pipeline, we demonstrate that the adaptive Sauvola method, in particular, significantly enhances model performance. The results show a 14.3 percentage point increase in mAP@0.5 (from 64.2% to 78.5%) compared to the unprocessed original image. The analysis reveals that adaptive binarization performs a fundamental feature selection function, removing low-level noise (blur, mold, texture) and highlighting the most relevant visual attributes for the model (shape and contour of musical elements).

Keywords: Optical Music Recognition; Digital Preservation; Historical Scores; Thresholding; Binarization; YOLOv11; Computer Vision; Feature Selection.

1. INTRODUÇÃO

A preservação do patrimônio cultural musical é um desafio de extrema relevância, uma vez que uma parcela significativa das composições originais de valor histórico e artístico permanece guardada exclusivamente em suporte físico, principalmente partituras em papel. Esses documentos são inerentemente vulneráveis à degradação ao longo do tempo, sofrendo os efeitos de agentes como umidade, fungos, dobras, rasgos e manchas. O Reconhecimento Ótico de Música é uma área da visão computacional focada em desenvolver soluções capazes de minerar e preservar símbolos musicais dentro de partituras, de uma forma que a lógica contida possa replicar e preservar as composições que hoje são sujeitas a degradação ao longo do tempo (Pacha; Hajić Jr.; Calvo-Zaragoza, 2018).

No entanto, a aplicação direta de modelos modernos de visão computacional e detecção de objetos a imagens de partituras degradadas apresenta limitações significativas. A contínua degradação de partituras introduz ruídos nas imagens digitalizadas, como manchas de fundo. Ademais, quando se trata de partituras em formato de papel, a textura, dobraduras ou variações de iluminação são entendidos como informações desnecessárias para modelos de visão computacional, prejudicando a performance destes. Isso leva modelos a falsos positivos, falsos negativos e uma redução geral na precisão da segmentação e classificação dos símbolos musicais (Pinto; Rebelo; Guedes, 2020). A complexidade inerente à notação musical, com sua diversidade de símbolos e relações espaciais, já representa um desafio substancial para sistemas de Reconhecimento Ótico de Música (Pacha; Hajić Jr.; Calvo-Zaragoza, 2018), sendo agravada por condições de degradação documental (Pinto *et al.*, 2011).

Com o intuito de minerar elementos contidos em partituras manuscritas preservados em formato de papel de forma otimizada, este trabalho propõe a aplicação de técnicas de pré-processamento às partituras, de uma forma que os atributos de interesse (elementos musicais) fiquem mais visíveis, e, posteriormente, avaliar o impacto de sua aplicação em métricas de performance de algoritmos de detecção de objetos. Com isso, formaliza-se um pipeline de Reconhecimento Ótico de Música onde a aplicação de algoritmos de limiarização se mostram como uma etapa crucial.

2. METODOLOGIA

A metodologia proposta para avaliar o impacto de técnicas de limiarização como seletor de atributos na performance de detecção

de elementos musicais é estruturada em um pipeline sequencial e comparativo. O núcleo do experimento consiste em aplicar dois algoritmos de binarização a um conjunto de imagens de partituras degradadas, para então aplicar um modelo de detecção, o YOLOv11, treinado para identificar elementos chave de partituras e medir a performance resultante em cada um dos casos. Abaixo, é ilustrado o pipeline do experimento.

Figura 1. Diagrama do pipeline do experimento. Desde a coleta do dataset, até a avaliação.

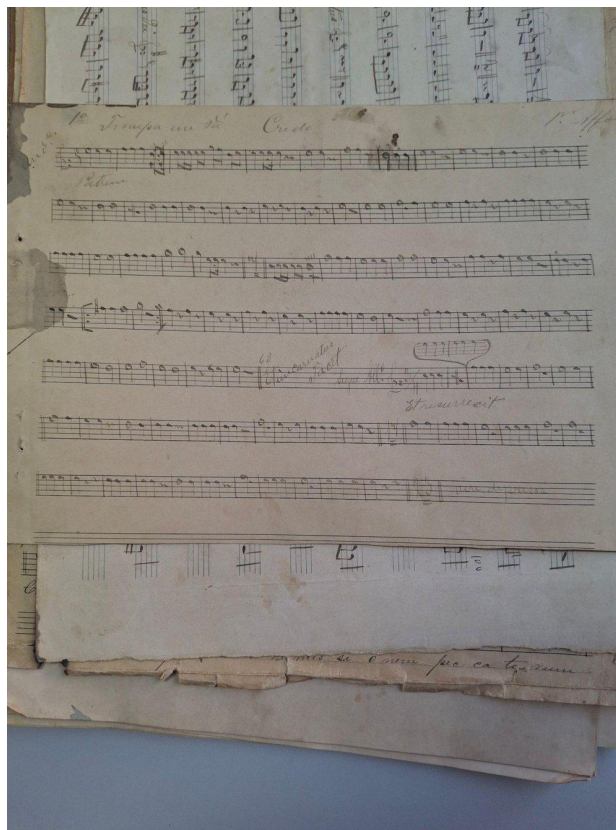


Fonte: Autores deste artigo (2025).

Conjunto de Dados

O conjunto de dados é composto por imagens digitalizadas de partituras históricas manuscritas, provenientes do acervo do Departamento de Música da Universidade Federal de São João del-Rei (UFSJ). As imagens apresentam variados graus de degradação física típica de documentos antigos. Não limitadas a manchas de umidade e mofo, amarelecimento e escurecimento do papel, dobras, rasgos leves e variações de contraste devido à digitalização não ideal. A Figura 2 exemplifica uma partitura degradada utilizada no estudo.

Figura 2. Exemplo de partitura degradada.

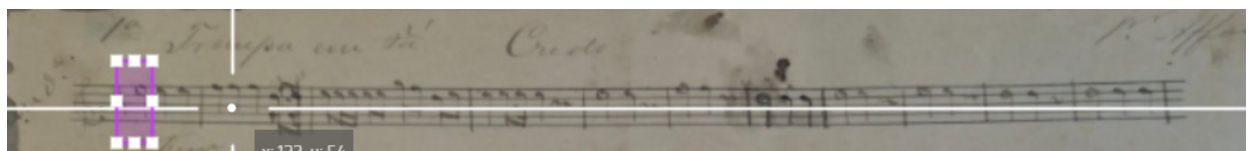


Fonte: Autores deste artigo (2025).

Especificamente, para este trabalho, o foco foi na detecção de elementos que definem a duração das notas. O modelo YOLOv11 (Ali; Zhang, 2024) foi treinado para reconhecer e classificar sete classes distintas relacionadas aos valores rítmicos: semibreve, mínima, semínima, colcheia, semicolcheia, fusa e semifusa.

Processo de Rotulagem

Figura 3. Demonstração do processo de rotulagem dos segmentos de partitura para o treinamento do modelo YOLO.



Fonte: Autores deste artigo (2025).

Fazendo uso da ferramenta Makesense.ai, cada elemento de interesse em cada linha de partitura é rotulada com a sua classe representativa. Gerando, assim, um documento de texto

correspondente a cada imagem no conjunto de dados. Cada linha de cada arquivo de texto possui as classes indicadas pelo rótulo, assim como as coordenadas da ocorrência a nível de pixel, a serem utilizados no processo de treinamento supervisionado do modelo YOLO a ser descrito posteriormente.

Técnicas de Limiarização

Cada imagem do conjunto de teste foi submetida a dois algoritmos de limiarização representativos: um global (Otsu) e um adaptativo local (Sauvola) (Fadlil, 2026). Uma versão da imagem sem qualquer pré-processamento (apenas redimensionada para a entrada da rede) foi mantida como base. O objetivo é avaliar qual técnica fornece o conjunto de atributos visuais mais “limpo” e discriminativo para o YOLOv11.

O algoritmo de Otsu é um método clássico que determina um único valor de limiar para toda a imagem, ideal para situações em que o fundo é uniforme e o histograma apresenta dois picos bem definidos. Entretanto, sua principal limitação surge ao processar documentos degradados ou com iluminação irregular, como partituras antigas. Nessas condições, por não considerar variações locais, o Otsu frequentemente seleciona características de forma equivocada, confundindo manchas ou degradações do fundo com elementos importantes, ou, inversamente, apagando símbolos musicais em áreas mais claras. Essa abordagem “cega” ao contexto tende a comprometer a preservação das informações relevantes.

Em contraste, o método de Sauvola foi concebido especificamente para a binarização de documentos, mostrando-se um seletor de características muito mais robusto para este contexto. Ele atua de

forma adaptativa, calculando um limiar distinto para cada pixel com base nas propriedades estatísticas de sua vizinhança. Seu mecanismo fundamental normaliza a variação local, suprimindo de maneira drástica o ruído e as texturas do fundo em regiões homogêneas, ao mesmo tempo que mantém uma resposta sensível nas áreas de alto contraste, onde geralmente se encontram os traços dos símbolos. Dessa forma, o Sauvola prioriza a forma e os detalhes dos elementos textuais ou gráficos relevantes, ignorando efetivamente as imperfeições do suporte. Os parâmetros do algoritmo, como o tamanho da janela de análise e os fatores de correção, foram ajustados empiricamente em um conjunto de validação para otimizar esse equilíbrio entre a preservação dos símbolos e a eliminação do ruído de fundo.

Modelo de Detecção de Objetos: YOLOv11

O YOLOv11 recebe as imagens de entrada (binarizada ou original), e os arquivos de anotações correspondentes a cada imagem. Ao longo do treinamento, é realizado o aprendizado dos atributos dos elementos anotados, assim como as coordenadas de suas caixas delimitadoras. Treinada a reconhecer os elementos de interesse e onde as procurar na imagem, a rede é utilizada para realizar previsões.

Inferência Controlada: O mesmo modelo treinado e congelado foi utilizado para realizar inferência nos três conjuntos de teste distintos, um utilizando o pré-processamento com o algoritmo Otsu, um com Sauvola e, por último, comparando-se com a imagem original. Dessa forma, qualquer variação de performance é atribuível exclusivamente à qualidade dos atributos de entrada gerados pela

etapa de limiarização, e não a diferenças no modelo ou em seu treinamento.

Protocolo Experimental e Avaliação

O experimento segue o seguinte fluxo:

1. Geração dos Conjuntos de Teste Processados: As imagens do conjunto de teste original foram processadas pelos dois algoritmos de limiarização, gerando três versões (original, limiarizada por Otsu e a limiarizada com Sauvola).
2. Inferência: O modelo YOLOv11 treinado foi executado em cada uma das três versões do conjunto de teste.
3. Métricas de Avaliação: As detecções foram comparadas com o subconjunto de validação utilizando métricas padrão:

Precision (Precisão): Proporção de detecções corretas entre todas as detecções realizadas. Mede a “pureza” dos atributos selecionados pela limiarização (menos falsos positivos).

Recall (Revocação): Proporção de objetos reais que foram corretamente detectados. Mede a capacidade da limiarização de preservar atributos dos símbolos verdadeiros (menos falsos negativos).

mAP@0.5 (mean Average Precision): Média da precisão para todas as classes, considerando uma Intersection over Union (IoU) de 0.5 como critério para detecção correta. É a métrica principal de agregação de desempenho.

F1-Score: Média harmônica entre Precision e Recall.

Análise Comparativa

Foi realizada uma análise quantitativa (tabelas, gráficos) e qualitativa (inspeção visual de casos críticos) para determinar qual técnica de limiarização, como seletor de atributos, proporciona o maior ganho global (mAP). Como cada algoritmo se comporta frente a diferentes tipos de degradação, com atenção especial às classes de duração (notas). A consistência dos benefícios entre as diferentes classes de elementos musicais, em particular entre os valores rítmicos.

3. RESULTADOS E DISCUSSÃO

Esta seção apresenta a análise quantitativa e qualitativa dos resultados obtidos com a aplicação do pipeline experimental. Os dados demonstram o impacto significativo da limiarização como etapa de seleção de atributos na performance do modelo YOLOv11, especialmente na detecção das classes de duração das notas. Ressalta-se, também, que a limiarização foi a única técnica de pré-processamento aplicada ao conjunto de dados.

Análise Quantitativa: Métricas Globais

A Tabela 1 resume as métricas médias de desempenho para o modelo YOLOv11 aplicado às três condições de entrada.

Tabela 1. Métricas comparativas de desempenho do YOLOv11 sob diferentes técnicas de limiarização (seleção de atributos).

Condição de Entrada	mAP@0.5	Precisão	Recall	F1-Score
---------------------	---------	----------	--------	----------

Original (Baseline)	64,2%	71,5%	68,3%	69,8%
+ Otsu (Global)	68,1% (+3,9 p.p.)	75,2%	70,9%	73,0%
+ Sauvola (Adaptativo)	78,5% (+14,3 p.p.)	83,4%	80,2%	81,8%

Fonte: Autores deste artigo (2025).

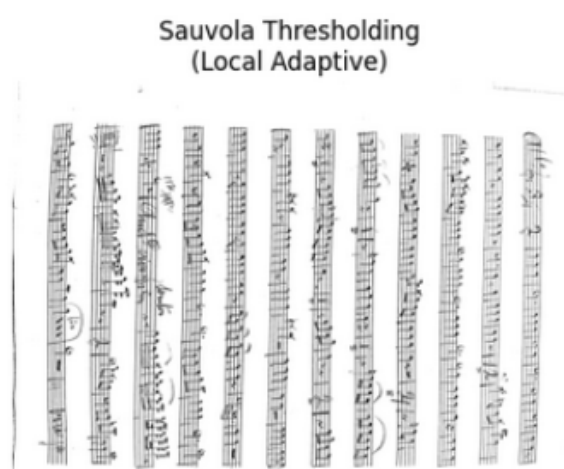
Os resultados são elucidativos e validam a hipótese central. O ganho na métrica principal, mAP@0.5, foi expressivo: o método de Sauvola proporcionou um aumento de 14,3 pontos percentuais, saltando de 64,2% para 78,5%. Trata-se de uma melhoria substancial na capacidade do YOLOv11 de localizar e classificar corretamente os objetos, especialmente por ele ser alimentado com atributos previamente selecionados e purificados pelo método adaptativo. Já o ganho modesto do Otsu (+3,9 p.p.) era esperado, uma vez que métodos globais costumam ser inadequados para cenários complexos com fundo não uniforme.

Quanto ao equilíbrio entre precisão e recall, o método de Sauvola novamente se destacou, alcançando a melhor harmonia entre os dois — com ganhos simétricos de cerca de 12 p.p. em cada um — o que se refletiu no maior F1-Score (81,8%). Isso mostra que a seleção de atributos feita pelo Sauvola é ao mesmo tempo eficiente e conservadora: remove ruídos que gerariam falsos positivos (aumentando a precisão) e preserva a integridade dos símbolos que poderiam levar a falsos negativos (aumentando o recall).

Análise Qualitativa: Casos de Sucesso e Falha

A análise visual (Figuras 4 e 5) ilustra o mecanismo de seleção de atributos em ação, com destaque para as notas no caso de sucesso (Sauvola), em que a Figura 4 ilustra o resultado da aplicação do algoritmo adaptativo local de limiarização sob uma partitura que apresentava desnível de iluminação ao longo da imagem.

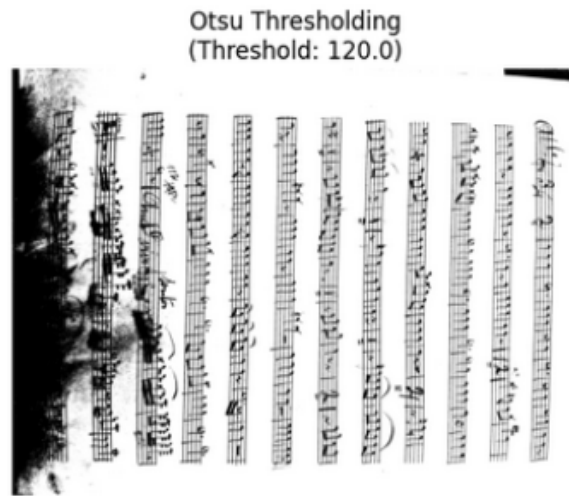
Figura 4. Demonstração de sucesso na habilidade do algoritmo Sauvola em encontrar limiares locais e realizar uma limpeza eficiente da imagem.



Fonte: Autores deste artigo (2025).

A Figura 5 mostra a aplicação do Otsu na mesma partitura. Nota-se a perda de informação: em regiões mais escuras (manchas), o limiar global escuro funde símbolos com o fundo, apagando atributos essenciais. Em regiões mais claras, o limiar pode ser muito baixo, introduzindo ainda mais ruído. O resultado é imagens binarizadas de qualidade inferior até mesmo quando comparadas com as originais.

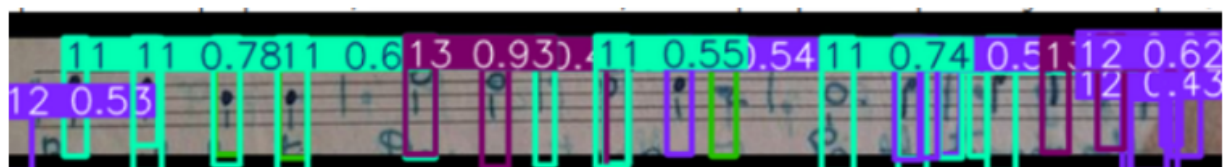
Figura 5. Demonstração da limitação do algoritmo Otsu em encontrar um limiar ideal em imagens com desnível de iluminação.



Fonte: Autores deste artigo (2025).

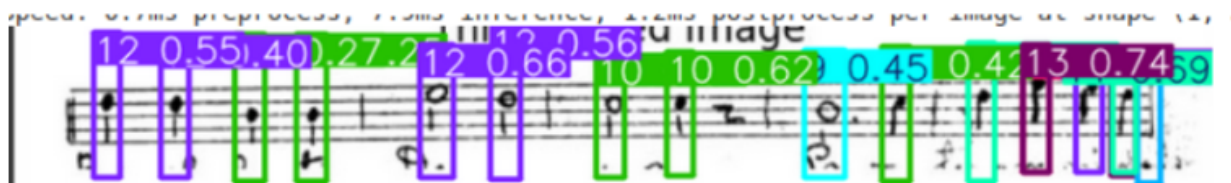
As Figuras 6a e 6b apresentam o comparativo direto das detecções realizadas pelo modelo treinado. Pode-se observar as detecções do YOLOv11 na imagem original degradada: diversas falsas detecções e várias notas não detectadas, especialmente as de menor duração. Na Figura 6b, observa-se o resultado na imagem processada com Sauvola: detecções limpas, precisas e abrangentes, validando o ganho qualitativo.

Figura 6a. Demonstração de detecção de elementos musicais em partitura sem pré-processamento.



Fonte: Autores deste artigo (2025).

Figura 6b. Demonstração de detecção de elementos musicais em partitura com pré-processamento.



Fonte: Autores deste artigo (2025).

4. CONSIDERAÇÕES FINAIS

O presente estudo investigou o impacto da aplicação de técnicas clássicas de limiarização como uma etapa de pré-processamento e, fundamentalmente, de seleção de atributos, para a detecção de elementos musicais em partituras históricas degradadas utilizando o modelo de última geração YOLOv11.

Através de um pipeline experimental controlado e comparativo, avaliou-se dois algoritmos representativos na sua capacidade de filtrar o ruído e apresentar ao YOLOv11 um conjunto de atributos visuais mais puro e discriminativo. O modelo YOLOv11 foi especificamente treinado para a detecção e classificação dos valores rítmicos fundamentais da notação ocidental: semibreve, mínima, semínima, colcheia, semicolcheia, fusa e semifusa.

Os resultados demonstram que ambas as técnicas melhoraram a performance, mas o método de Sauvola destacou-se de forma marcante, elevando a métrica principal mAP@0.5 em 14,3 pontos percentuais (de 64,2% para 78,5%) em relação à linha de base sem processamento.

A discussão dos resultados à luz do referencial teórico permitiu concluir que a eficácia do Sauvola reside em sua função dupla como um seletor de atributos inteligente: (1) supressor ativo de atributos irrelevantes (variações de cor, texturas, sombras) e (2) realçador e

isolador das atributos geometricamente definidas dos símbolos musicais (contornos, formas). Esta pré-seleção simplifica radicalmente o problema de aprendizagem do YOLOv11, permitindo que a rede concentre seus parâmetros na modelagem da relação entre forma e semântica musical, em vez de desperdiçar capacidade na tentativa de ignorar o ruído. A análise detalhada das classes de duração demonstrou que os benefícios foram mais pronunciados para os símbolos de traços mais finos e complexos (semicolcheia, fusa, semifusa), que são justamente os mais suscetíveis à degradação.

Este estudo demonstra que a implementação de uma etapa de pré-processamento na frente de um modelo baseado em dados (YOLOv11) constitui uma estratégia viável para aumentar a eficácia de modelos de detecção de objetos. Esta abordagem híbrida é valiosa para a digitalização de acervos históricos, onde a qualidade da imagem é frequentemente abaixo do ideal e a anotação de grandes conjuntos de dados para treinar modelos extremamente robustos contra ruídos é inviável.

5. TRABALHOS FUTUROS

O presente trabalho limitou o seu escopo para a análise do impacto de técnicas de limiarização baseadas em algoritmos na performance do modelo YOLO como ferramenta de detecção de elementos musicais em partituras manuscritas. Trabalhos futuros possuem duas frentes de interesse para expansão dos experimentos.

Primeiramente, realizar experimentos em técnicas de limiarização baseadas em dados, em que modelos de aprendizado profundo são treinados para a decisão de ativação ou não de cada pixel em uma

imagem no processo de limiarização. Sendo uma solução baseada em dados, a qualidade de sua implementação estará sujeita à qualidade do dataset de treinamento.

A segunda frente de interesse para a continuidade da linha de pesquisa é o teste da performance de modelos de detecção de objetos além do YOLO. Algoritmos de detecção de objetos diferentes possuem peculiaridades quanto à performance em diferentes casos de uso. Então, segue que testes de diferentes algoritmos possam encontrar performances variadas para a detecção de elementos musicais em partituras manuscritas.

REFERÊNCIAS BIBLIOGRÁFICAS

AFZAL, M. Z.; PASTOR-PELLICER, J.; SHAFAIT, F.; BREUEL, T. M.; DENGEL, A.; LIWICKI, M. Document image binarization using LSTM: a sequence learning approach. In: INTERNATIONAL CONFERENCE ON DOCUMENT ANALYSIS AND RECOGNITION (ICDAR), 2015. *Proceedings [...]*. [S. l.: s. n.], 2015.

ALI, M. L.; ZHANG, Z. The YOLO framework: a comprehensive review of evolution, applications, and benchmarks in object detection. *Computers*, v. 13, n. 12, p. 336, 2024. DOI: <https://doi.org/10.3390/computers13120336>.

BHARGAVA, N.; BANSAL, A. A comprehensive review on image preprocessing techniques in document images. *Multimedia Tools and Applications*, v. 80, n. 18, p. 27867-27904, 2021.

BISWAS, R.; SARKHEL, S.; ROY, S. K.; PAL, U. Transdocunet: a transformer-based unet architecture for degraded document image

binarization. In: ACM SYMPOSIUM ON DOCUMENT ENGINEERING (DOCENG), 2024. *Proceedings [...]*. [S. l.: s. n.], 2024.

CALVO-ZARAGOZA, J.; RIZO, D. End-to-end neural optical music recognition of monophonic scores. *Applied Sciences*, v. 8, n. 4, p. 606, 2018.

FADLIL, A. Performance evaluation of Otsu and Sauvola thresholding for structured document binarization. *Scientific Journal of Informatics*, v. 13, n. 1, p. 189-200, 2026. DOI: <https://doi.org/10.15294/sji.v13i1.40245>.

FUJINAGA, I.; HANKINSON, A. Music document analysis and recognition. In: *Springer Handbook of Systematic Musicology*. Berlin, Heidelberg: Springer, 2018. p. 889-905.

GATOS, B.; PRATIKAKIS, I.; PERANTONIS, S. J. Adaptive degraded document image binarization. *Pattern Recognition*, v. 39, n. 3, p. 317-327, 2006.

JINDAL, H.; KUMAR, M.; TOMAR, A.; MALIK, A. Degraded document image binarization using novel background estimation technique. In: INTERNATIONAL CONFERENCE FOR CONVERGENCE IN TECHNOLOGY (I2CT), 6., 2021. *Proceedings [...]*. [S. l.: s. n.], 2021. p. 1-8.

LU, S.; TAN, C. L. Thresholding of badly illuminated document images through photometric correction. In: INTERNATIONAL CONFERENCE ON DOCUMENT ANALYSIS AND PATTERN RECOGNITION, 2007. *Proceedings [...]*. [S. l.: s. n.], 2007.

OTSU, N. A threshold selection method from gray-level histograms. *IEEE Transactions on Systems, Man and Cybernetics*, v. 9, n. 1, p. 62-

66, 1979.

PACHA, A.; HAJIĆ JR., J.; CALVO-ZARAGOZA, J. A baseline for general music object detection with deep learning. *Applied Sciences*, v. 8, n. 9, p. 1488, 2018.

PINTO, T.; REBELO, A.; GIRALDI, G.; CARDOSO, J. S. Music score binarization based on domain knowledge. In: IBERIAN CONFERENCE ON PATTERN RECOGNITION AND IMAGE ANALYSIS (IBPRIA), 2011. *Pattern Recognition and Image Analysis*. Berlin: Springer, 2011. (Lecture Notes in Computer Science, v. 6669). p. 700-708.

PINTO, T.; REBELO, A.; GUEDES, C. Robust staffline removal in handwritten music scores. *International Journal on Document Analysis and Recognition (IJDAR)*, v. 23, n. 3, p. 205-218, 2020.

REBELO, A.; CAPELA, G.; CARDOSO, J. S. Optical recognition of music symbols: a comparative study. *International Journal on Document Analysis and Recognition (IJDAR)*, v. 15, n. 1, p. 19-31, 2012.

REDMON, J.; DIVVALA, S.; GIRSHICK, R.; FARHADI, A. You only look once: unified, real-time object detection. In: IEEE CONFERENCE ON COMPUTER VISION AND PATTERN RECOGNITION (CVPR), 2016. *Proceedings [...]*. [S. l.: s. n.], 2016.

REIS, D.; KUPEC, J.; BARTOS, V.; TATAR, M. Real-time flying object detection with YOLOv8. *arXiv preprint*, arXiv:2305.09972, 2023.

SAUVOLA, J.; PIETIKÄINEN, M. Adaptive document image binarization. *Pattern Recognition*, v. 33, n. 2, p. 225-236, 2000.

SU, B.; LU, S.; TAN, C. L. Binarization of historical document images using the local maximum and minimum. In: DOCUMENT RECOGNITION AND RETRIEVAL (DPR), 2010. *Proceedings [...]*. [S. l.: s. n.], 2010.

TUGGENER, L.; ELEZI, I.; SCHMIDHUBER, J.; STADELMANN, T. Deep watershed detector for music object recognition. *arXiv preprint*, arXiv:1805.10548, 2018.

VAN DER WEL, E.; ULLRICH, K. Optical music recognition with convolutional sequence-to-sequence models. In: *ISMIR*, 2017.

¹ Universidade Federal de São Carlos (UFSCar), São Carlos, SP, Brasil.
E-mail: [acesse o artigo original para visualizar o e-mail](#)

² Universidade Estadual de Campinas (Unicamp), Campinas, SP, Brasil. E-mail: [acesse o artigo original para visualizar o e-mail](#)

³ Universidade Federal de São João del-Rei (UFSJ), São João del-Rei, MG, Brasil. E-mail: [acesse o artigo original para visualizar o e-mail](#)

⁴ Universidade Federal de São João del-Rei (UFSJ), São João del-Rei, MG, Brasil. E-mail: [acesse o artigo original para visualizar o e-mail](#)