# Um Sistema de Validação de Imagens de Documentos Pessoais Utilizando Detecção de Objetos

**Lucas S. Fernandes[1], Francisco Igor da Silva Lima[1], Tácio Soares Aguiar[1], Rodrigo da Silva Freitas[1], Gabriel Campos de Oliveira[1], José Gilvan Rodrigues Maia[1], Paulo Antonio Leal Rego[1]**

[1]Departamento de Computação
Universidade Federal do Ceará (UFC) – Fortaleza, CE – Brasil

{lucasdesousafernandes,igorlima_es,tacioaguiar,

rodrigosf,gabrielcdo}@alu.ufc.br,

gilvan@virtual.ufc.br,paulo@dc.ufc.br

***Abstract.*** *Image-based identity document analysis and recognition have been used in various applications and contexts, bringing benefits to society. Many processes are manually executed and would benefit from automating the validation of information contained in these documents. This work presents a novel automatic system for personal document validation. It extracts and recognizes relevant information using a state-of-the-art object detection model and optical character recognition. Using birth and marriage certificates as object of study, private dataset of real documents is used for training the network and testing the system. The information recognition achieved 75% accuracy for name recognition, 82% for registration, and 65% for both together. Considering at least 80% similarity, our method achieved 94% accuracy for name recognition, 93% for registration, and 91% for both together. Thus, these results indicate that the proposed method is promising.*

***Resumo.*** *A análise e o reconhecimento de documentos de identidade baseados em imagens têm sido usados em várias aplicações e contextos, trazendo benefícios para a sociedade. Muitos processos são realizados manualmente e se beneficiariam da automatização da validação de informação contida nesses documentos. Este trabalho apresenta um novo sistema automático de validação de documentos pessoais. Ele extrai e reconhece informações relevantes usando um modelo de detecção de objetos de última geração e reconhecimento óptico de caracteres. Utilizando certidões de nascimento e casamento como objeto de estudo, um conjunto de dados privado de documentos reais é usado para treinar a rede e testar o sistema. O reconhecimento de informações atingiu 75% de precisão para reconhecimento de nomes, 82% para matrículas e 65% para ambos juntos. Considerando pelo menos 80% de similaridade, nosso método alcançou 94% de precisão para o reconhecimento de nomes, 93% para o registro e 91% para ambos juntos. Portanto, esses resultados indicam que o sistema proposto é promissor.*

## 1. Introdução

Atualmente, muitas tarefas relacionadas a documentos pessoais são realizadas manualmente, consumindo tempo e recursos. Entretanto, com o avanço da tecnologia, é razoável

assumir que os modernos sistemas computadorizados especializados com métodos de reconhecimento de padrões podem ser capazes de efetivamente compreender a estrutura dos documentos [Tang et al. 1991]. Nesse sentido, surgem nos últimos anos modelos de Inteligência Artificial em Documentos para automaticamente classificar, ler e obter informação de documentos. Esses modelos englobam técnicas diversas de Processamento de Imagens e de Aprendizagem de Máquina. Esta última facilitou em grande escala o trabalho humano, sendo utilizada em diversas aplicações, como buscas na web, filtragem de conteúdo, assistentes virtuais, sistemas de recomendação de conteúdo, etc [LeCun et al. 2015].

A certidão de nascimento é o primeiro e mais importante desses documentos de identidade para os cidadãos brasileiros[1]. Somente com essa certidão uma pessoa existe oficialmente para o Estado e a sociedade, e pode emitir outros documentos de identidade, tais como o registro geral, o título de eleitor, o cartão de contribuinte individual, a carteira de trabalho, etc. Além disso, existe a certidão de casamento, um documento cujo conteúdo é extraído do acordo matrimonial elaborado em um livro depositado aos cuidados de um cartório de registro civil e substitui a certidão de nascimento. A Figura 1 mostra exemplos de certidões de nascimento e casamento brasileiras, anônimas para proteger informações privadas.


(a) (b)

**Figura 1. Exemplos de certidões brasileiras de (a) nascimento e (b) casamento.**

Neste contexto, o processamento manual de certidões pode ser uma tarefa custosa, propensa a erros, e, como os documentos de identidade (ID), pode se beneficiar da automação. Os sistemas de reconhecimento de IDs [Bulatov et al. 2022, Polevoy et al. 2022] são amplamente utilizados para obter e verificar as informações pessoais dos usuários em muitas aplicações. Em vários processos cotidianos, os cidadãos precisam apresentar documentos durante serviços presenciais ou virtuais, e.g., para abrir uma conta bancária, solicitar benefícios sociais e solicitar outros documentos como RG e Passaporte, entre vários outros cenários. De acordo com Bulatov et. al. [Bulatov et al. 2022], uma grande parte da literatura é direcionada à criação de sistemas de análise de IDs com foco específico em um subconjunto de tipos de documentos ou em um modo particular de obtenção de imagem. No entanto, há uma crescente demanda por reconhecimento de documentos de identidade a partir de uma grande variedade de fontes de imagem, tais como varreduras, fotos ou molduras de vídeo, bem como em uma variedade de *layouts* e condições de captura praticamente incontrolada [Polevoy et al. 2022].

[1]https://www.planalto.gov.br/ccivil_03/leis/l6015compilada.htm

A tarefa de processamento de documentos de identidade é complicada por muitas questões, como problemas no momento da digitalização do documento (luminosidade, ângulo de rotação do documento, má qualidade de digitalização, etc) e documentos com baixa qualidade de leitura (rasuras, amolgadelas, arranhões, furos, etc). No caso de certidões, eles podem variar muito dependendo do cartório onde foram emitidos, dificultando abordagens mais gerais e baseadas em modelos.

O processo de análise e extração de dados em documentos é desafiador, como definido acima, mas traz vários benefícios para os cidadãos e para os colaboradores ou funcionários que processam manualmente os documentos. Alguns benefícios são:

- Reduz o número de rejeições de falsos positivos e automatiza um processo manual custoso e demorado;
- Maior eficiência no processo de validação de informações, reduzindo pequenos erros de digitação;
- Permite a busca de antecedentes, multas, processos judiciais e outras informações vitais de segurança para o processo em questão, utilizando as informações extraídas dos documentos; e
- Redução do tempo de serviço, automatizando a extração e verificação de dados, reduzindo assim as filas de serviço e aumentando a escalabilidade do processo.

## 2. Trabalhos relacionados

Nesta seção, são apresentados vários trabalhos com objetivos similares, contendo análise e reconhecimento de documentos pessoais, mais especificamente de identidade [Bulatov et al. 2022, Castelblanco et al. 2020, Xu and Wu 2018, Wu et al. 2019]. É importante destacar as técnicas utilizadas nestes trabalhos e seus respectivos resultados. Além disso, são apresentados datasets contendo documentos de identidade brasileiros gerados artificialmente.

Bulatov et al [Bulatov et al. 2022] analisaram as modernas técnicas de reconhecimento de documentos e propuseram uma estrutura unificada para análise e reconhecimento de documentos de identidade. As tarefas estudadas consistem no préprocessamento, preparação, extração e reconhecimento de cararacteres. Para isso, os autores trabalharam em técnicas de detecção, localização e processamento de modelos para destacar informações do documento e avaliar seus resultados.

Castelblanco et al. [Castelblanco et al. 2020] apresentaram uma ferramenta de verificação de identidade, tanto de classificação como de legitimidade dos documentos. No desenvolvimento, os autores cuidaram de vários aspectos que poderiam criar dificuldades para a tarefa, incluindo posição, perspectiva, ângulo, brilho e qualidade de imagem dos documentos. O processo é baseado em dois módulos que correspondem à aquisição de documentos e sua verificação. Para executar essas tarefas, os autores utilizam algoritmos de aprendizagem de máquina e processamento de imagem. Os resultados são satisfatórios e promissores, o que justifica a implementação futura com outros tipos de documentos.

Jianxing Xu et al. [Xu and Wu 2018] expuseram uma versão melhorada da abordagem tradicional para reconhecimento de documentos de identidade. Os autores dividiram o processo em três etapas: correção da angularidade, localização da região do texto e segmentação e OCR. Para a correção da rotação de documentos, eles utilizam a transformada de Hough. Para a detecção de texto, os autores utilizam classificadores que utilizam

atributos Haar da imagem para definir regiões. Finalmente, um modelo de rede neural é treinado a partir de um conjunto de dados de caracteres chineses e ingleses e números arábicos para o OCR. Os resultados se provam satisfatórios.

Xing Wu et al. [Wu et al. 2019] apresentaram uma abordagem para autenticação de identidade em dispositivos móveis, a fim de evitar roubo de identidade. Isto é feito usando um método de verificação facial a partir de redes neurais convolucionais, e um método para reconhecer o texto presente na identificação com modelos de redes neurais. Os resultados provam ser satisfatórios e mais precisos do que os métodos mais modernos.

Tratando de documentos pessoais brasileiros, o trabalho de Soares et. al. [Álysson Soares et al. 2020] apresenta um base de imagens de documentos pessoais brasileiros, em especial CNHs, CPFs e RGs. As imagens, em frente e verso dos documentos, são geradas artificialmente com base em imagens de documentos reais anonimizadas. O dataset contém 28800 imagens no total, 3600 para cada classe, e tem como objetivo abordar os seguintes problemas relacionados a visão computacional: classificação de imagens de documentos, segmentação de região de textos e reconhecimento óptico de caracteres. O dataset gerado não tem marcações dos campos relevantes, dificultando sua utilização para avaliar métodos de extração de informações.

Lopes et. al. [Lopes Junior et al. 2021] propõem um dataset para a competição ICDAR 2021 focado em segmentação de componentes em fotos de documentos. Os dados foram coletados de documentos brasileiros - RGs, CNHs e CPFs, tendo suas informações pessoais substituídas. Para a competição, foram considerados três desafios: segmentação das fronteiras dos documentos, segmentação das zonas de texto e segmentação das assinaturas. Cada desafio tem 15000 imagens de treino e 5000 imagens de teste. 16 equipes submeteram resultados, com os melhores utilizando segmentação através de redes neurais.

## 3. Proposta

Nesta seção, é descrita a proposta de sistema automático de validação de documentos pessoais, composto por: (A) detecção de documentos, (B) extração de informações e (C) validação das informações. Os documentos considerados são, por exemplo, certidões de nascimento e casamento, registros gerais (RGs) , cadastros de pessoas físicas (CPFs) e carteiras nacionais de habilitação (CNHs). A Figura 2 mostra a *pipeline* de detecção e extração de informações. Com as informações extraídas, é possível cruzar as informações submetidas pelo usuário do sistema com as informações obtidas pelo processamento descrito para, então, serem validadas.

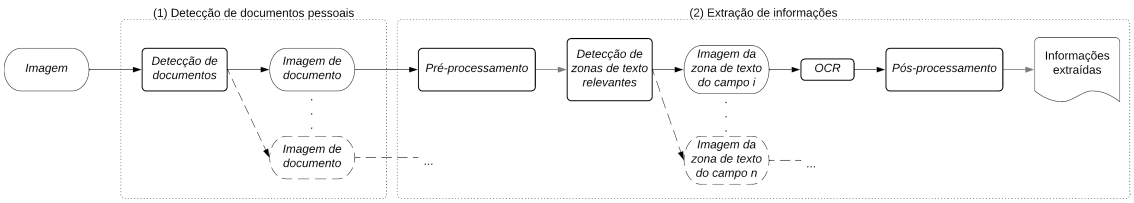


**Figura 2. *Pipeline* de extração de informações de documentos pessoais**

A base dessa pesquisa são modelos de detecção de objetos através de aprendizagem profunda, sendo uma escolha robusta para processamento de imagens em tempo

real. As redes neurais possuem grande capacidade de generalização, sendo mais robustas que métodos tradicionais de processamento digital de imagens, e os métodos de detecção de objetos em específico pode ser rápida e eficiente em termos computacionais, em comparação com métodos mais complexos como segmentação de objetos pixel a pixel. Deste modo, a escolha por esse tipo de modelo é clara.

### 3.1. Detecção de documentos

Dada uma imagem contendo um documento pessoal, o primeiro passo para a extração de informações é a identificação, categorização e isolamento do documento contido na imagem. Esta etapa é realizada por um método de detecção de objetos utilizando aprendizagem profunda, sem a necessidade de pré-processamento ou entradas manuais. A imagem poderia ser, por exemplo, um escaneamento ou uma foto tirada de um *smartphone*, contendo um ou mais documentos em destaque. O modelo é, então, treinado em um amplo conjunto de imagens previamente marcadas. O resultado é, portanto, uma imagem do documento em si isolado e a qual categoria ele pertence - certidão de nascimento ou casamento, RG, CNH ou CPF, para cada documento contido na imagem. Estas podem, então, ser utilizadas na próxima etapa.

### 3.2. Extração de informações

Com o documento devidamente categorizado e isolado, as informações relevantes podem ser extraídas através de um método de detecção de objetos. O modelo é treinado em um amplo conjunto de imagens previamente marcadas. Este identifica e isola a zona de texto na imagem correspondente a um campo determinado do documento como, por exemplo, o número de inscrição do Registro Geral. O texto contido na região de interesse pode, então, ser lido através de um método de OCR. Enfim, os pares de campo e seu respectivo texto lido são as informações extraídas de cada documento.

### 3.3. Validação das informações

Nesse cenário, o sistema recebe como entrada a imagem do documento para extrair informações e um formulário preenchido pelo usuário contendo as informações necessárias a serem validadas pelo sistema. Com as informações extraídas das imagens, pode-se comparar com as informações de entrada e automaticamente validar a informação submetida como correta ou incorreta, evitando um trabalho de verificação manual e já rejeitando falsas entradas.

## 4. Resultados preliminares

Esta seção descreve os resultados preliminares dos experimentos para extração de informação em documentos pessoais, utilizando certidões de nascimento e casamento como objeto de estudo.

As bases de imagens foram inicialmente compostas por arquivos digitalizados em formato PDF, fornecidos pela Secretaria de Segurança Pública e Defesa Social do Estado do Ceará em parceria com UFC. Em seguida, esses arquivos foram convertidos em imagens através do pacote pdf2image em Python [2].

[2]https://pypi.org/project/pdf2image/

## 4.1. Dataset

O dataset foi dividido em 1063 imagens para treinamento, 300 para validação, e 200 para teste. Dentro das imagens do dataset de teste, 120 imagens tiveram o seu texto anotado para avaliação dos resultados de leitura das informações. As certidões no dataset são diversos. Existem documentos mais antigos sem número de matrícula e formatos que variam pelo cartório que os emitiu. Isso torna mais difícil generalizar como extrair informações a partir destas imagens, sendo difícil utilizar *template-matching*, por exemplo. A Figura 3 mostra alguns exemplos das imagens no dataset.


(a) (b) (c) (d)

**Figura 3. Exemplos do dataset de certidões. De (a) e (b) são certidões de nascimento e de (c) a (d) são certidões de casamento.**

## 4.2. Treinamento e testes

A *pipeline* do método, ilustrada na Figura 4, é composta por: (A) detecção de informações relevantes (nome e matrícula) em imagens de certidões e (B) reconhecimento das informações através de OCR. No primeiro passo, a informação é detectada através de YOLOv8, versão mais recente do modelo, para isolar os textos relevantes. Finalmente, a informação detectada é reconhecida usando Tesseract-OCR e uma heurística é aplicada como pós-processamento para amenizar possíveis erros.

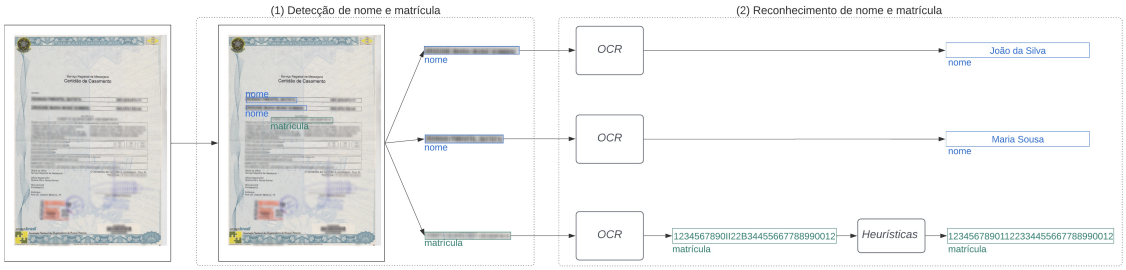


**Figura 4. Pipeline da metodologia proposta de extração de informação contextualizada em um exemplo hipotético.**

O equipamento utilizado foi um computador composto por um AMD Ryzen 5 5600 como CPU, uma NVIDIA GeForce RTX 3060ti como GPU e Windows 11 como sistema operacional. A biblioteca Ultralytics foi utilizada para treinamento e inferência da rede YOLO, com PyTorch como base.

### 4.2.1. Detecção de informações

A abordagem proposta considera o número de matrícula e nome(s) como informações essenciais a serem extraídas da imagem da certidão. O número de matrícula existe nas certidões a partir de 2009 e é a fonte de muitas informações sobre a certidão. Ele contém qual livro, folha e termo do cartório foi registrada e se é uma certidão de nascimento ou casamento. Em experiências anteriores, foi testada a extração dessas informações apenas aplicando OCR em todo o documento e procurando por um número de 32 dígitos e se contém o nome de entrada, mas os resultados não foram interessantes, falhando em muitos casos. Assim, utilizar uma arquitetura de detecção de objetos, ou seja, YOLOv8 para detectar e isolar essas informações em imagens de certidões é a alternativa estudada neste trabalho. Assim, a YOLO detecta as informações relevantes na imagem de entrada, e então elas são isoladas em imagens separadas que serão usadas na próxima etapa. Este detector pode ser substituído em futuras implementações.

Para o treinamento do detector de informações no dataset mencionado, primeiro, inspirado nas técnicas *human-in-the-loop* [Russakovsky et al. 2015, Yu et al. 2015], foi utilizado um dataset menor de 200 imagens em que os quatro cantos de todos os nomes e matrículas foram marcados manualmente e treinado o detector YOLO no mesmo. Depois, o dataset completo de treinamento foi *pseudo-marcado* pelo próprio detector, pelo que cada anotação foi validada manualmente e revista quando necessário. Enfim, a rede foi treinada neste dataset.

Quanto ao treinamento em si, a rede usada foi a versão YOLOv8s. Uma abordagem usando a validação cruzada k-fold para mostrar a capacidade de generalização do método foi adotada, mais especificamente 5-fold. Cada treinamento levou aproximadamente uma hora. A Figura 5 mostra a *loss* da rede diminuindo e a precisão média (mAP) aumentando ao longo das épocas no última *fold*. Os resultados da validação cruzada são mostrados na Tabela 1.

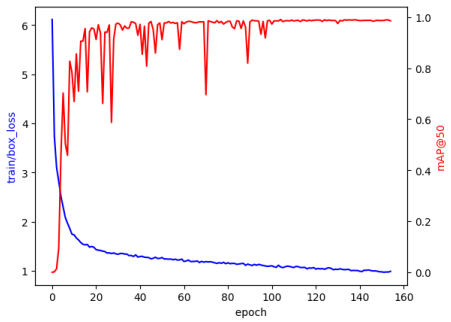


**Figura 5. Treinamento da rede YOLO, loss e mAP da validação ao longo das iterações**

Os resultados dos datasets de validação e teste são mostrados na Tabela 2. Com uma mAP@50 de 97.7%, precisão de 0.960 e sensibilidade de 0.924, pode-se considerar que o detector de informação provê resultados robustos.

**Tabela 1. Resultados da validação cruzada para detecção de informações usando YOLOv8**

| | **Precisão** | **Sensibilidade** | **mAP@50** |
|---|---|---|---|
| Ambos | $0.981 \mp 0.010$ | $0.969 \mp 0.009$ | $0.989 \mp 0.003$ |
| Matrícula | $0.989 \mp 0.0064$ | $0.979 \mp 0.017$ | $0.994 \mp 0.001$ |
| Nome | $0.973 \mp 0.015$ | $0.9576 \mp 0.00$ | $0.982 \mp 0.005$ |

**Tabela 2. Resultados de teste para detecção de informações usando YOLOv8**

| | **Precisão** | **Sensibilidade** | **mAP@50** |
|---|---|---|---|
| Ambos | 0.960 | 0.924 | 0.977 |
| Matrícula | 0.953 | 0.933 | 0.980 |
| Nome | 0.966 | 0.916 | 0.974 |

### 4.2.2. Leitura da informação

Dadas as imagens isoladas do número de matrícula e dos nomes (um nas certidões de nascimento ou dois nas certidões de casamento) e considerando uma margem de 10% em cada lado para garantir que todo o texto seja capturado, esta próxima etapa realiza o OCR. Mais especificamente, a implementação atual utiliza Tesseract-OCR [Smith 2007] para reconhecer o texto em cada imagem, tratando-os como uma única linha de texto. Em seguida, o texto é processado para limpar símbolos, acentos, espaços e caixa rebaixada. Uma heurística simples é aplicada no texto de matrícula, substituindo erros comuns que são reconhecidos como letras por dígitos visualmente semelhantes, como mostra a Tabela 3.

**Tabela 3. Letras a serem substituídas como dígitos em textos de números de matrícula reconhecidos**

| letra | dígito |
|---|---|
| A / a | 4 |
| B / b | 8 |
| C / c | 0 |
| I / i | 1 |
| L / l | 1 |
| O / o | 0 |

Nesta etapa final, os textos das matrículas e nomes de 120 imagens do dataset dos testes foram marcados. Algumas delas não continham matrícula e foram ignoradas. Então, o texto reconhecido pelo OCR pôde ser comparado com o texto anotado. A Tabela 4 mostra os resultados do OCR, considerando um resultado como correto se o texto for idêntico e incorreto caso contrário. Além disso, mostra os resultados considerados corretos se o texto for pelo menos 80% semelhante, usando similaridade de Jairo. O primeiro caso de uso, considerando textos idênticos, é mais adequado para a filtragem de textos digitados incorretamente. O segundo, considerando a semelhança entre textos, é mais tolerante, identificando quando o documento está parcialmente correto e evitando entradas erradas. Além disso, para efeitos de comparação, uma validação básica foi implementada

utilizando OCR nas imagens e procurando o texto anotado e, como mostra a Tabela, o método proposta já se mostra superior a essa *baseline*.

**Tabela 4. Resultados da leitura (OCR) utilizando Tesseract-OCR**

| | **Acurácia / Sensibilidade** | |
|---|---|---|
| | **Método proposto** | Baseline |
| Completo | 0.65 | 0.31 |
| Nome | 0.75 | 0.50 |
| Matrícula | 0.82 | 0.38 |
| Completo (80% de similaridade) | 0.91 | |
| Nome (80% de similaridade) | 0.94 | |
| Matrícula (80% de similaridade) | 0.93 | |

A Figura 6 mostra alguns exemplos de resultados em certidões de nascimento e casamento do dataset. As caixas de delimitação azul e verde mostram nomes e registros detectados e reconhecidos, respectivamente. O texto foi anonimizado para fins de privacidade, mostrando apenas alguns poucos caracteres para ilustrar os resultados.


(a) (b) (c)

**Figura 6. Alguns exemplos de resultados da abordagem estudada**

## 5. Discussão e Conclusão

Em resumo, este trabalho tratou de um sistema de validação de documentos pessoais através de extração de informações, propondo uma arquitetura utilizando detecção de objetos com redes neurais. Foi utilizado o modelo de detecção de objetos estado-da-arte YOLOv8 para detectar informação relevante em imagens de certidões de nascimento e casamento e Tesseract-OCR para reconhecer o texto de cada informação detectada. Deste modo, foi alcançado um resultado de 65% de acurácia considerando o documento completo, 75% para nomes e 82% para matrículas. Os resultados preliminares são promissores e mostram o potencial deste trabalho em contribuir para a área de processamento de documentos pessoais. Nas próximas etapas desse trabalho, pretende-se expandir a proposta e experimentos para outros tipos de documentos brasileiros, como registro geral e carteira nacional de habilitação. Além disso, um estudo é visado para analisar possíveis técnicas de processamento para propor modelos de detecção e OCR mais especializados

no problema abordado, alcançando melhores resultados, e propor uma base de imagens pública para comparação dos resultados.

## Referências


Bulatov, K., Bezmaternykh, P., Nikolaev, D., and Arlazarov, V. (2022). Towards a unified framework for identity documents analysis and recognition. *Computer Optics*, 46(3):436–454.

Castelblanco, A., Solano, J., Lopez, C., Rivera, E., Tengana, L., and Ochoa, M. (2020). Machine learning techniques for identity document verification in uncontrolled environments: A case study. In *Mexican Conference on Pattern Recognition*, pages 271–281. Springer.

LeCun, Y., Bengio, Y., and Hinton, G. (2015). Deep learning. *nature*, 521(7553):436.

Lopes Junior, C. A., das Neves Junior, R. B., Bezerra, B. L., Toselli, A. H., and Impedovo, D. (2021). Icdar 2021 competition on components segmentation task of document photos. In *Document Analysis and Recognition–ICDAR 2021: 16th International Conference, Lausanne, Switzerland, September 5–10, 2021, Proceedings, Part IV 16*, pages 678–692. Springer.

Polevoy, D. V., Sigareva, I. V., Ershova, D. M., Arlazarov, V. V., Nikolaev, D. P., Ming, Z., Luqman, M. M., and Burie, J.-C. (2022). Document liveness challenge dataset (dlc-2021). *Journal of Imaging*, 8(7):181.

Russakovsky, O., Li, L.-J., and Fei-Fei, L. (2015). Best of both worlds: human-machine collaboration for object annotation. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 2121–2131.

Smith, R. (2007). An overview of the tesseract ocr engine. In *Ninth international conference on document analysis and recognition (ICDAR 2007)*, volume 2, pages 629–633. IEEE.

Tang, Y. Y., Suen, C. Y., Yan, C. D., and Cheriet, M. (1991). Documents analysis and understanding: a brief survey.

Wu, X., Xu, J., Wang, J., Li, Y., Li, W., and Guo, Y. (2019). Identity authentication on mobile devices using face verification and id image recognition. *Procedia Computer Science*, 162:932–939.

Xu, J. and Wu, X. (2018). A system to localize and recognize texts in oriented id card images. In *2018 IEEE International Conference on Progress in Informatics and Computing (PIC)*, pages 149–153. IEEE.

Yu, F., Seff, A., Zhang, Y., Song, S., Funkhouser, T., and Xiao, J. (2015). Lsun: Construction of a large-scale image dataset using deep learning with humans in the loop. *arXiv preprint arXiv:1506.03365*.

Álysson Soares, das Neves Junior, R., and Bezerra, B. (2020). Bid dataset: a challenge dataset for document processing tasks. In *Anais Estendidos do XXXIII Conference on Graphics, Patterns and Images*, pages 143–146, Porto Alegre, RS, Brasil. SBC.