



UNIVERSIDADE FEDERAL DO RIO GRANDE - FURG
CENTRO DE CIÊNCIAS COMPUTACIONAIS
PROGRAMA DE PÓS-GRADUAÇÃO EM COMPUTAÇÃO
CURSO DE MESTRADO EM ENGENHARIA DE COMPUTAÇÃO

Dissertação de Mestrado

**Avaliação de Qualidade em Imagens de Documentos
Técnicos usando Fusão de Métricas**

Lucas Viana da Silva

Dissertação de Mestrado apresentada ao Programa de Pós-Graduação em Computação da Universidade Federal do Rio Grande - FURG, como requisito parcial para a obtenção do grau de Mestre em Engenharia de Computação

Orientador: Prof. Dr. Paulo Lilles Jorge Drews-Jr

Rio Grande, 2024

Ficha Catalográfica

S586a Silva, Lucas Viana da.

Avaliação de qualidade em imagens de documentos técnicos usando fusão de métricas / Lucas Viana da Silva. – 2024.
57 f.

Dissertação (mestrado) – Universidade Federal do Rio Grande – FURG, Programa de Pós-Graduação em Computação, Rio Grande/RS, 2024.

Orientador: Dr. Paulo Lilles Jorge Drews-Jr.

1. OCR 2. Qualidade 3. *Machine learning* 4. IQA I. Drews-Jr, Paulo Lilles Jorge II. Título.

CDU 004

Catálogo na Fonte: Bibliotecário José Paulo dos Santos CRB 10/2344

ATA DE SESSÃO DE DEFESA DE DISSERTAÇÃO DE MESTRADO

Ata nº 06/2024

Aos vinte e sete dias do mês de setembro de dois mil e vinte e quatro, às quatorze horas e trinta minutos, ocorreu a sessão de Defesa de Dissertação de Mestrado de LUCAS VIANA DA SILVA, que apresentou a dissertação intitulada **“Avaliação de Qualidade em Imagens de Documentos Técnicos usando Fusão de Métricas”**, realizada sob a orientação do Prof. Dr. Paulo Lilles Jorge Drews Junior. A banca examinadora foi constituída pelos Prof. Dr. Eduardo Nunes Borges (FURG) e Prof. Dr. Giancarlo Lucca (UCPEL), sob a presidência do orientador. Após a apresentação do trabalho, a banca arguiu o candidato e, a seguir, deliberou pela

- (X) aprovação da Dissertação
- () aprovação da Dissertação, sugerindo modificações no texto
- () reprovação da Dissertação

Rio Grande, 27 de setembro de 2024.



C=BR, O=ICP-Brasil, OU=VideoConferência, OU=07215063000104, OU=Secretaria da Receita Federal do Brasil - RFB, OU=RFB e-CPF A3, OU=(em branco), CN=EDUARDO NUNES BORGES,00057035067 Rio Grande - RS 2024.11.04 14:31:12-03'00'

Prof. Dr. Eduardo Nunes Borges (participação remota)

GIANCARLO

LUCCA:01808620062

Assinado de forma digital por
GIANCARLO LUCCA:01808620062
Dados: 2024.09.27 16:18:44 -03'00'

Prof. Dr. Giancarlo Lucca (participação remota)

Assinado de forma digital por Paulo
Lilles Jorge Drews Junior:00124140009
Dados: 2024.09.27 21:14:51 +02'00'

Prof. Dr. Paulo Lilles Jorge Drews Junior (participação remota)
Orientador

RESUMO

SILVA, Lucas Viana da. **Avaliação de Qualidade em Imagens de Documentos Técnicos usando Fusão de Métricas**. 2024. 55 f. Dissertação (Mestrado) – Programa de Pós-Graduação em Computação. Universidade Federal do Rio Grande - FURG, Rio Grande.

Reconhecimento óptico de caracteres (OCR) e restauração de documentos são duas aplicações em que fazer a avaliação da qualidade da imagem do documento (DIQA) é necessária para entender os problemas que podem surgir ao tentar processar um documento específico. Documentos técnicos, contendo formas complexas como tabelas e carimbos/assinaturas, apresentam comportamento diferente na saída de um OCR dependendo de qual componente do documento está sendo analisado, fazendo com que ocorra problemas ao utilizar os modelos de DIQA atuais. Este estudo analisará estas diferenças de comportamento e proporá um método para a utilização de DIQA neste tipo de documento, separando os componentes do mesmo e tratando cada um deles de forma separada. Os resultados obtidos mostram uma melhora perto de 14% em comparação com técnicas do estado da arte atual para documentos que incluem formas complexas como tabelas e carimbos, adicionando uma nova ferramenta relativamente simples para tarefas de DIQA.

Palavras-chave: OCR, Qualidade, Machine learning, IQA.

ABSTRACT

SILVA, Lucas Viana da. **Quality Assessment in Technical Document Images using Metrics Fusion**. 2024. 55 f. Dissertação (Mestrado) – Programa de Pós-Graduação em Computação. Universidade Federal do Rio Grande - FURG, Rio Grande.

Optical Character Recognition (OCR) and Document Information Retrieval are two applications where doing Document Image Quality Assessment (DIQA) is necessary to predict problems that can arise when trying to process a specific document. Technical documents, containing complex forms such as tables, stamps and signatures, have different behavior in the output of an OCR depending on which component of the document is being analyzed, causing problems when using current DIQA models. This study will analyze these differences in behavior and propose a method for using DIQA in this type of document, separating its components and treating each one separately. The results obtained show close to a 14% improvement compared to current state-of-the-art techniques on documents that include complex shapes such as tables and stamps, proving a great tool for a niche usage in the DIQA segment.

Keywords: OCR, Quality, Machine learning, IQA.

LISTA DE FIGURAS

Figura 1	Fluxograma de processo OCR utilizando a análise de qualidade . . .	9
Figura 2	Um exemplo de documento técnico [15].	11
Figura 3	Exemplo de tabela [32]	17
Figura 4	Exemplo de carimbo [32]	17
Figura 5	Exemplo de página do dataset StaVer [32]	19
Figura 6	Exemplo de página do dataset FUNSD [15]	20
Figura 7	Exemplo de página sem utilização de <i>data augmentation</i>	22
Figura 8	Exemplo de página utilizando <i>Inkbleed</i> com tamanho de <i>kernel</i> (5,5).	23
Figura 9	Exemplo de página utilizando <i>Folding</i>	24
Figura 10	Exemplo de página utilizando <i>DirtyDrum</i> de forma vertical.	25
Figura 11	Exemplo de página utilizando <i>DirtyDrum</i> de forma horizontal.	25
Figura 12	Exemplo de página utilizando <i>DirtyDrum</i> de forma vertical e horizontal.	25
Figura 13	Exemplo de página utilizando <i>Letterpress</i>	26
Figura 14	Exemplo de página utilizando compressão JPEG com 50% de qualidade.	27
Figura 15	Exemplo de página utilizando compressão JPEG com 10% de qualidade.	27
Figura 16	Exemplo de página utilizando compressão JPEG com 1% de qualidade.	27
Figura 17	Exemplo de página utilizando <i>Scribbles</i>	28
Figura 18	Treinamento de uma rede DIQA.	31
Figura 19	Execução de uma rede DIQA.	32
Figura 20	Documento original. <i>Recognition Score</i> : 0,90	43
Figura 21	Segmento "tabela" do documento. <i>Recognition Score</i> : 0,85	43
Figura 22	Segmento "carimbo" do documento. <i>Recognition Score</i> : 0,62	43
Figura 23	Segmento "base" do documento. <i>Recognition Score</i> : 0,95	44

LISTA DE TABELAS

Tabela 1	Resultados de acurácia do TesseractOCR - FUNSD	28
Tabela 2	Resultados de acurácia do PaddleOCR - FUNSD	29
Tabela 3	Correlação de técnicas por componente - Par Tesseract e StaVer - LCC	36
Tabela 4	Correlação de técnicas por componente - Par Tesseract e StaVer - SROCC	36
Tabela 5	Resultados de qualidade DIQA - Par Tesseract e Staver	37
Tabela 6	Correlação de técnicas por componente - Par Paddle e StaVer - LCC .	37
Tabela 7	Correlação de técnicas por componente - Par Paddle e StaVer - SROCC	37
Tabela 8	Resultados de qualidade DIQA - Par Paddle e Staver	38
Tabela 9	Correlação de técnicas por componente - Par Tesseract e FUNSD - LCC	39
Tabela 10	Correlação de técnicas por componente - Par Tesseract e FUNSD - SROCC	39
Tabela 11	Resultados de qualidade DIQA - Par Tesseract e FUNSD	39
Tabela 12	Correlação de técnicas por componente - Par Paddle e FUNSD - LCC	40
Tabela 13	Correlação de técnicas por componente - Par Paddle e FUNSD - SROCC	40
Tabela 14	Resultados de qualidade DIQA - Par Paddle e FUNSD	40
Tabela 15	Correlação entre predição-valor real - <i>Cross dataset</i>	41
Tabela 16	Resultados de qualidade DIQA - Par Tesseract e Staver - Sem <i>data augmentation</i>	42

LISTA DE ABREVIATURAS E SIGLAS

IQA	Image Quality Assessment
OCR	Optical Character Recognition
DIQA	Document Image Quality Assessment
NR-IQA	No-Reference Image Quality Assessment
CER	Character Error Rate
SVR	Support Vector Regression
CNN	Convolutional Neural Network
RNN	Recurrent Neural Network
PDF	Portable Document Format
JPEG	Joint Photographic Experts Group
WER	Word Error Rate
N-WER	Normalized Word Error Rate
LSTM	Long Short-Term Memory
MSE	Mean Square Error
LCC	Linear Correlation Coefficient
SROCC	Spearman Rank Order Correlation Coefficient
IoU	Intersection-over-Union
CRNN	Convolutional Recurrent Neural Network
SSIM	Structural Similarity Index Measure

SUMÁRIO

1	Introdução	9
1.1	Definição de problema	11
1.2	Objetivos	11
1.3	Estrutura do texto	12
2	Trabalhos relacionados	14
2.1	Conclusão	16
3	Metodologia	17
3.1	Dataset	18
3.2	Data Augmentation	21
3.2.1	<i>Inkbleed</i>	23
3.2.2	<i>Folding</i>	23
3.2.3	<i>DirtyDrum</i>	25
3.2.4	<i>Letterpress</i>	25
3.2.5	Artefatos JPEG	25
3.2.6	<i>Scribbles</i>	25
3.3	Motores OCR	27
3.3.1	Tesseract	27
3.3.2	Paddle	29
3.4	Métricas de qualidade de OCR	29
3.5	Métodos de DIQA	30
3.6	Métrica de saída do OCR	32
3.6.1	Linear Correlation Coefficient	32
3.6.2	Spearman Rank Order Correlation Coefficient	33
3.7	Fusão de valores de qualidade	33
4	Resultados	35
4.1	StaVer	35
4.1.1	Tesseract	35
4.1.2	Paddle	36
4.2	FUNSD	38
4.2.1	Tesseract	38
4.2.2	Paddle	39
4.3	Estudos complementares	41
4.3.1	<i>Cross Dataset</i>	41
4.3.2	<i>Data Augmentation</i>	41

4.4	Exemplo qualitativo	42
5	Conclusão	45
	Referências	47

ANEXOS

A	Hipóteses	52
A.1	Validação	52
A.1.1	Hipótese 1	52
A.1.2	Hipótese 2	53
A.1.3	Hipótese 3	53

1 INTRODUÇÃO

Com a utilização crescente de sistemas digitais se faz necessário a tradução de documentos que não foram criados de forma digital, e portanto não tem anotado o texto dentro do mesmo, em documentos que podem ser reconhecidos e manipulados através de sistemas de computação, como sistemas de busca de texto. Para isso é feito o processo de OCR nos documentos, visando a extração do texto em uma imagem específica. Embora os sistemas de OCR tiveram uma melhora significativa nos últimos anos, com adventos como redes Transformer, ainda é possível notar que em algumas imagens, especialmente as imagens com problemas de qualidade, a extração não consegue ser feita de forma satisfatória. Com isso em mente foram criadas técnicas de *Document Image Quality Assessment* (DIQA), para tentar medir a qualidade de um documento antes de fazer o processamento pelo OCR, fazendo com que seja possível adaptar o processo de extração de texto dependendo da qualidade predita, adicionando etapas como pré-processamento da imagem ou mudando a rede utilizada durante o OCR. O fluxograma deste processo pode ser visto na Figura 1. As técnicas de DIQA atuais focam em imagens com texto manuscrito ou texto em mundo real, não sendo adaptadas para textos de documentos técnicos escaneados, que é um tipo de documento razoavelmente comum nos tempos atuais, consistindo desde plantas de construção civil até artigos científicos.

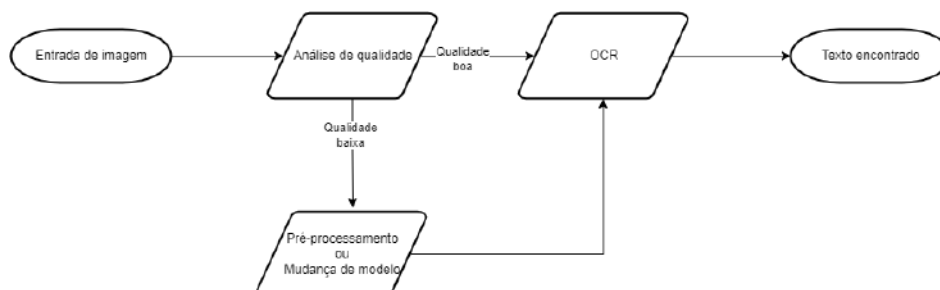


Figura 1: Fluxograma de processo OCR utilizando a análise de qualidade

Image Quality Assessment(IQA) é um método de regressão utilizado em diversas áreas, desde imagens médicas até documentos de texto, para tentar analisar a qualidade da imagem com relação a algum processo, por exemplo o quão bom a qualidade de uma imagem de radiologia é para o reconhecimento de um tumor, sendo que existem alguns

subtipos de IQA com relação a qualidade de documentos (DIQA). A primeira subdivisão tem haver com qual métrica de qualidade está sendo avaliada, podendo ser subjetiva ou objetiva. Uma métrica subjetiva [21] tende a utilizar a percepção humana como qualidade, levantando os dados de qualidade de cada imagem a ser analisada através de um estudo realizado com pessoas, onde as mesmas determinam um número em uma escala definida para um determinado conjunto de imagens. Métricas subjetivas não são muito utilizadas nos métodos de DIQA, visto que processos de OCR podem facilmente errar uma imagem com texto de reconhecimento fácil para uma pessoa, como texto com uma linha riscada por cima do mesmo.

Métricas objetivas se dividem em dois tipos, sendo eles métodos utilizando a saída do OCR com algum processo de aprendizado e métodos que só utilizam as informações da imagem, sem utilização de aprendizado de máquina. Métodos que não utilizam aprendizado de máquina [43] [4] são criadas a partir da extração de características das imagens e a aplicação de filtros nas características para tentar estimar um valor de *score* pré-definido através de uma fórmula matemática.

Métodos que tentam prever a qualidade do OCR através de aprendizado de máquina utilizam a comparação entre a saída do OCR e o texto anotado da imagem, normalmente analisando o mesmo com métricas como *Character Error Rate* (CER) utilizando a distância de Levenshtein, CLEval [2] ou ZoneMapAlt [19] para obter a acurácia da detecção e reconhecimento do texto feito pelo OCR. Estes métodos utilizam métodos clássicos em conjunto com a extração manual de características como SVR [37] [46] ou redes neurais como CNN [29] [26] [30] e RNN [27]. Estes métodos então tentam fazer uma regressão da uma métrica de saída do OCR, que é tomada como métrica de qualidade, a partir de uma imagem dada.

Vale ressaltar também que IQA pode ser dividido em dois tipos:

- IQA com referência completa (FR-IQA): onde o processo compara uma imagem de entrada com uma imagem conhecida e com qualidade "perfeita". Normalmente utiliza algoritmos para comparar as duas imagens [6]. Alguns algoritmos que são utilizados para este problema são *structural similarity index measure* (SSIM) [45] e *universal quality index* (UQI)[44]. Este tipo não é visto no ramo de DIQA, porque os processos que utilizam de técnicas de DIQA não tem uma imagem de referência para comparar a imagem que está tendo a qualidade inferida. Normalmente essas técnicas são mais relevantes em processos que precisam calibrar equipamentos como câmeras digitais ou para notar deterioração ao longo dos equipamentos de captura de imagem utilizados em um local estático.
- IQA sem referência (NR-IQA) ou *blind IQA*: onde existe uma imagem de entrada e nenhuma referência/comparação sobre a qualidade da mesma. É o tipo com maior número de estudos, sendo que o trabalho desenvolvido neste estudo, e em todos

aparecem em uma página do documento. Isto será feito através de um método de divisão e conquista, segmentando cada elemento complexo na página e aplicando a melhor técnica de DIQA em cada um para depois juntar o resultado dos mesmos com o resultado do resto da página, considerado como elemento "base" neste trabalho. A hipótese inicial que gerou a criação deste método foi de que os sistemas OCR atuais se comportam de forma diferente em elementos complexos, fazendo com que técnicas de DIQA tenham dificuldade de atribuir um valor correto para a qualidade do documento total, sendo essa hipótese validada no anexo A.1.2.

Alguns objetivos secundários do projeto são:

- Experimentar técnicas que se provam efetivas em outras áreas de IQA e analisar seu comportamento em DIQA [40, 31].
- Analisar correlação de valores de qualidade com diversas métricas para quantificar o resultado de um sistema OCR [24, 3, 36].
- Explorar métodos de fusão para o conjunto de divisão e conquista [33].
- Utilizar mais de um sistema OCR para verificar a generalidade do método.
- Utilizar diversos tipos de documentos escaneados, contendo e não contendo elementos complexos, para verificar a acurácia do método em outros tipos de documentos que não são o foco principal do trabalho.

Para a realização destes objetivos será necessário levantar uma base de dados que contenha elementos de tabela e carimbo. Com essa base será possível fazer o treinamento de redes e testar técnicas de DIQA na página inteira e nos elementos isolados, fazendo com que seja possível isolar o processo para obter uma predição de qualidade melhor para cada componente. Por último, essas predições de qualidade para cada componente de um documento precisam ser fundidas em uma predição geral, dando então um valor de qualidade final para aquele documento.

1.3 Estrutura do texto

Este trabalho é estruturado em 5 capítulos:

- Capítulo 1(Introdução): introduz o problema de DIQA em documentos técnicos, a estrutura e os objetivos do trabalho proposto.
- Capítulo 2(Trabalhos relacionados): explora o estado da arte para problemas parecidos ao tema do projeto.
- Capítulo 3(Metodologia): explica o método utilizado para lidar com o problema explicado durante a introdução.

- Capítulo 4(Resultados preliminares): demonstra os resultados obtidos com o método explicado na seção anterior.
- Capítulo 5(Conclusão): recapitula os resultados obtidos na seção anterior e considera áreas que poderiam ser estendidas no mesmo para um futuro projeto.

2 TRABALHOS RELACIONADOS

Como o foco do trabalho está em DIQA sem referência focado em documentos técnicos no texto abaixo serão destacados alguns artigos do estado da arte focando especificamente em DIQA. Vale ressaltar a maioria dos trabalhos de DIQA não tem o foco somente em documentos técnicos, muitas vezes misturando diversos tipos de documento em uma base de dados e focando no processo de DIQA em si, ao contrário deste trabalho que foca na aplicação de DIQA em um tipo de documento específico.

Li et al. [25] é um exemplo de DIQA com visão computacional clássica, isto é, sem a utilização de técnicas de aprendizado de máquina. A técnica consiste em utilizar um filtro de Sobel [18] sobre áreas de texto detectadas na imagem para tentar detectar e atribuir um valor a degradação nas bordas dos caracteres encontrados na área. Estes valores obtidos são agregados em um único valor através do desvio padrão dos mesmos. Como a métrica de qualidade escolhida no artigo não é uma métrica de saída de um sistema OCR é necessário correlacionar a mesma a acurácia do OCR para que seja possível analisar o resultado obtido e como ele se comporta quanto métrica de qualidade, isto é feito utilizando as métricas de correlação *Spearman Rank Order Correlation Coefficient* (SROCC) e *Linear Correlation Coefficient* (LCC). Os resultados obtidos obtêm bons resultados quando comparados com o estado da arte, mas o método contém algumas peculiaridades, como o fato do mesmo medir somente a qualidade do reconhecimento do texto e não medir como é a qualidade da imagem para a detecção do texto, outra parte importante em sistemas OCR. Também vale destacar que o banco de dados utilizado só contém imagens de livros, sem considerar outras formas de imagens, como documentos técnicos e imagens de texto na vida real.

Peng et al. [37] é um exemplo de DIQA utilizando aprendizado de máquina clássico, fazendo um processo de DIQA sem referência utilizando duas características extraídas com algoritmos de visão clássica, gradiente de borda e a razão entre a altura e a largura do texto detectado, combinando as mesmas com uma *Support Vector Regression*(SVR) para conseguir aprender a relação entre a métrica de qualidade de um processo OCR sobre as imagens, no caso a *Word Error Rate*(WER), e as características extraídas. Este processo é feito em imagens degradadas a partir de uma mudança de foco na câmera que

obteve as imagens da base de dados, sem utilizar nenhum pós-processamento para criar imagens degradadas por outros efeitos. A acurácia final do processo foi medida utilizando *Mean Square Error* (MSE), uma métrica obtida pela diferença entre a saída do regressor SVR treinado e a saída real do sistema OCR. Embora este método obteve resultados razoáveis o mesmo ainda tem alguns problemas quanto ao tipo de degradação das imagens utilizadas, analisando somente imagens com degradação devido a problemas de foco na câmera de captura, e também sobre o tipo de métrica de qualidade a ser aprendida, onde WER causa um valor desproporcional a ser atribuído em erros de severidade diferente, por exemplo o valor de WER se um sistema OCR trocar a palavra "casa" por "Casa" é o mesmo de se o sistema trocar "casa" por "C5s", porque a métrica WER só olha se uma palavra foi encontrada de forma correta ou não, onde a maioria dos estudos envolvendo análise de acurácia de sistemas OCR atuais utilizam métricas mais robustas como CER para representar a qualidade do processo OCR por causa de sua capacidade de considerar os erros de OCR no nível de caráter.

Lu and Dooms [29] é um exemplo de DIQA utilizando aprendizado profundo, especificamente redes neurais convolucionais. Este artigo foca no desempenho de redes neurais convolucionais profundas e no problema de que bases de dados para treino de modelos DIQA são pequenas, o que dificulta este tipo de rede a obter um bom resultado. Utilizando o mecanismo de transferência de aprendizado de uma rede já treinada em outras tarefas de visão computacional, neste caso a AlexNet [20], o artigo faz o *fine-tuning* da rede, reconfigurando a mesma para a regressão dos valores de acurácia de OCR, utilizando métricas de correlação entre a acurácia real do OCR e a acurácia predita pelo modelo. Este método obteve bons resultados quando comparado com o estado da arte, mas a sua base de dados de teste e treino não utiliza nenhum documento técnico, a acurácia do OCR só mede o reconhecimento de texto (CER) e não mede a detecção do texto para um sistema OCR e só existe um tipo de degradação nos dados.

Li et al. [27] utiliza uma rede neural recorrente (RNN), marcando o primeiro uso de uma rede que utiliza mecanismos de atenção para fazer a correlação entre a imagem de entrada e a qualidade da mesma para um processo OCR. Ele utiliza uma rede RNN em conjunto com uma métrica atribuída a cada imagem do conjunto de treinamento, neste caso a taxa de erros por caractere (CER), para treinar um modelo de regressão capaz de prever a qualidade de uma imagem para o processamento de um sistema OCR. Este artigo é interessante pois o mesmo trás o primeiro *dataset* focado em documentos técnicos, no caso recibos, e demonstra que embora o método seja capaz de prever a qualidade de imagens em documentos normais, com 95% de correlação pelo coeficiente LCC, o mesmo só consegue alcançar 81% de correlação em documentos mais complexos, como recibos.

Jin et al. [17] demonstra um método para a área de IQA em geral que mostra uma avaliação de qualidade que propõe dividir as sub-áreas de uma imagem em um dos três ti-

pos de blocos: liso, borda e textura. Para cada tipo sub-área da imagem é feita a avaliação de qualidade com base na melhor métrica inferida para o tipo da mesma, com métricas diferentes dependendo do tipo de bloco e do tipo de degradação observada em uma imagem. Após ser feita a avaliação para cada sub-área é necessário a fusão de todos os valores encontrados a partir de uma média aritmética. Jin et al. [16] expande essa pesquisa na parte de fusão de métricas, analisando também média ponderada, média truncada, mediana e média geométrica. Também é expandido com métricas utilizadas em problemas de IQA gerais chamadas de *Visual Attention Models* que utilizam macro-regiões que englobam múltiplas sub-áreas. Os resultados desse artigo mostram um aumento de acurácia significativo ao utilizar uma abordagem de divisão-e-conquista com múltiplas métricas, melhorando em quase 7% no coeficiente SROCC em comparação a utilização de uma única métrica para a imagem inteira.

Para os métodos baseados de DIQA que envolvem aprendizado descritos acima, vale ressaltar que os mesmos tentam prever a acurácia do OCR em uma imagem. Visto que OCR é o processo principal de análise de imagens de documentos, o desempenho de qualquer aplicação depende diretamente do desempenho do OCR. Assim, vários métodos de DIQA na literatura consideram a precisão do OCR, como uma métrica, para quantificar a qualidade das imagens dos documentos.

2.1 Conclusão

A motivação para este método saiu de uma análise feita em documentos técnicos na subseção A.1.2, onde foi observado que os métodos atuais de DIQA se mostram competentes em documentos sem forma complexa, porém os mesmos tem uma qualidade menor quando utilizado em documentos com carimbos e tabelas. Durante essa análise também foi possível notar que os métodos atuais de DIQA conseguem lidar com os componentes de forma isolada, isto é, aplicando a técnica somente na área da imagem que contém a tabela ou o carimbo. Para tratar este problema o trabalho a seguir irá propor um método que utiliza de partes dos métodos anteriores, mas também adiciona a ideia de análise de qualidade de segmentos e fusão de qualidade de componentes específicos a documentos técnicos.

3 METODOLOGIA

O processo de análise de qualidade de imagem deste trabalho é definido para documentos técnicos, contendo tabelas, formulários e carimbos, não-manuscritos e escaneados. O intuito do projeto é melhorar a previsão da qualidade através da separação desses documentos através de um processo de divisão e conquista, dividindo a imagem em componentes distintos, aplicando uma técnica de DIQA para prever a qualidade do componente e, ao final, juntar todos as previsões de qualidade de componentes em um valor só para obter o valor de qualidade do documento. Os componentes utilizados são:

- Tabelas(exemplo na Figura 3).
- Carimbos(exemplo na Figura 4).
- Base: o resto da página onde não existem carimbos nem tabelas.

Pos.	Anzahl	Beschreibung	Einzelpreis	Gesamt preis
	1	Versandkosten	1.21	1,21 €
1	13	WLAN AP ZisLink AP5431	69.75	906,75 €
2	28	Optical Mouse PXC-233	15.49	433,72 €
3	14	Radiowecker	13.86	194,04 €
Gesamtsumme				1535,72 €
inkl. 19% MwSt				245,20 €

Figura 3: Exemplo de tabela [32]

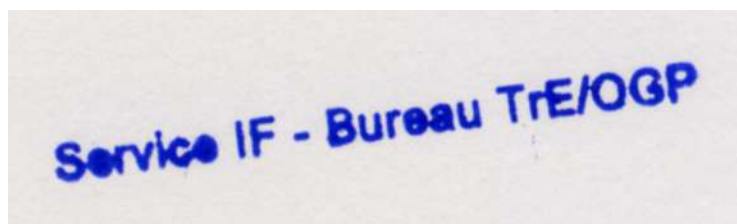


Figura 4: Exemplo de carimbo [32]

Este processo de dividir e conquistar foi feito com base nas hipóteses de que os sistemas OCR tendem a variar sua acurácia por componente dependendo da qualidade, isto é,

com uma degradação piorando muito um componente e não afetando tanto os outros. A segunda hipótese que foi levantada é que, por não contarem com essa mistura de elementos em uma página e serem treinados prioritariamente em bases de dados que não contém esse tipo de imagem, os modelos de DIQA atuais não teriam uma boa performance no tipo de documento em que o projeto foca. A última hipótese levantada considera que os sistemas de detecção destes componentes são robustos o suficiente a degradação da qualidade da imagem, em comparação ao OCR que não conseguiu lidar tão bem com as mesmas. Essas hipóteses foram verificadas no anexo A.

O projeto então é dado em duas etapas, sendo a primeira o treinamento/utilização de técnicas de DIQA em cada componente para analisar qual o melhor resultado em cada um dos mesmos. A segunda etapa consiste na fusão dos valores de qualidade obtidos para a comparação com os outros métodos de DIQA atuais. As subseções seguintes apresentaram os motores de OCR a serem utilizados nos experimentos, o levantamento dos *datasets* que serão utilizados, os métodos de DIQA que a serem analisados e os métodos de como essa fusão de valores de qualidade será feita.

3.1 Dataset

Para a seleção da base de dados alguns critérios foram levantados:

- Precisa conter documentos técnicos.
- Precisa conter diversos tipos de degradação ou imagens sem degradação onde possa ser aplicado *data augmentation*, para testar o processo em diversas situações de qualidade.
- Documentos precisam conter o texto anotado.

Além destes critérios também foi dado prioridade a *datasets* que contêm carimbos e tabelas, para que o método de agregação proposto no trabalho possa ser utilizado. Com esses critérios e preferências foram utilizadas para os experimentos duas bases de dados: StaVer [32] e FUNSD [15].

A primeira é uma base de dados de documentos técnicos escaneadas e não-manuscritas em alemão com foco em carimbos. Foi escolhido por existir os três elementos, tabelas, carimbos e texto normal, embora a área de onde se encontra cada tabela, o texto que se encontra em cada carimbo e o texto da página inteira teve que ser anotada manualmente. Um exemplo de página da mesma pode ser visto na Figura 5. A base contém 400 páginas com três níveis de qualidade diferente durante o escaneamento, com 200 DPI, 300 DPI e 600 DPI, totalizando em 1200 páginas na base completa. A extensão das anotações manuais feitas na base StaVer para este projeto podem ser vistas no link.

A segunda base de dados consiste de documentos técnicos escaneados e não-manuscritos em inglês com foco em formulários. A mesma foi escolhida para analisar

Service IF - Bureau THE/OGP

Matthias Werner, Bachstraße 184, 09626 Krempe

Jan Weber
Jahnstraße 94
83280 Stadt Wehlen

Ihre Bestellung vom
22. Juni 1997

LIQUIDATION

Sehr geehrter Herr Weber,

Bitte überweisen Sie den folgenden Rechnungsbeitrag innerhalb von 28 Tagen auf das unten angegebene Konto.

Anzahl	Beschreibung	Einzelpreis	Gesamtpreis
1	Versandkosten	9.19	9,19 €
26	50m Kabeltrommel, Überspannungsschutz	25.99	675,74 €
27	6x Steckdose	5.99	161,73 €
18	4x Steckdose	4.99	89,82 €
Gesamtsumme			936,48 €
inkl. 19% MwSt			149,52 €

Es grüßt Sie
Matthias Werner

free culture research conference

Matthias Werner
Bachstraße 184
09626 Krempe

Telefon: (039 3785)035-7010
Fax: (039 3785)435-7010
E-Mail: (039 3785)435-7010

Datum
15. November 2010

Gesellschafter:
Matthias Werner
UStId-Nr
DE 6913369676

Bankverbindung:
Konto 118 781
Sparkasse Kasselburg
BLZ 950 092 92

Geschäftsführer:
Matthias Werner
Gerichtstand:
Sankt Andreasberg

Figura 5: Exemplo de página do dataset StaVer [32]

o método proposto em uma base de dados onde os componentes de carimbo e tabela não são tão prevalentes em todas as páginas do dataset. Embora todas as imagens tem o mesmo tratamento durante sua digitalização, sendo todas escaneadas a 100 DPI, a inclusão desta base de dados também foi considerada pela qualidade de algumas imagens na mesma, com ruídos naturais de tinta durante a impressão e rotações naturais das páginas entre cada escaneamento das mesmas. Um exemplo de página da mesma pode ser vista na Figura 6. A quantidade total de imagens na mesma é de 199 páginas escaneadas com a anotação de texto para cada uma, necessitando também a anotação manual de tabelas, sem a necessidade anotação de carimbos pois o mesmo não sem encontra entre as páginas. A extensão das anotações manuais feitas na base FUNSD para este projeto podem ser vistas no link.

10675

S **STOUT INDUSTRIES, INC.**
6425 W. FLORISSANT AVE., ST. LOUIS, MO. 63136 • (314) 385-2280

PROPOSAL					
TO: <u>Lorillard Corporation</u> ADDRESS: <u>666 Fifth Avenue</u> CITY: <u>New York</u> STATE: <u>New York 10103</u> Attn: Mr. Robert Kennedy	FOR: <u>Metal "Pack" Plaque</u> DATE: <u>October 16, 1987</u> Mr. A. D. Steinberg YOUR REPRESENTATIVE				
<i>It is our pleasure to propose the following:</i>					
ITEM: <u>Harley Davidson Metal Plaque</u> SIZE: <u>17 1/4 x 2 3/4"</u> MATERIAL: STEEL ☐ ALUMINUM ☒ OTHER ☐ GAUGE: <u>.025"</u> COLORS: <u>Transparent gold, opaque black, white and orange</u> BASE COLOR: <u>Aluminum</u> SINGLE FACE ☒ DOUBLE FACE ☐ HOLES: YES ☒ NO ☐ NUMBER OF: <u>4</u> CORNERS: ROUND ☐ SQUARE ☒ ANGLE ☐ CUT TO SHAPE ☐ EDGES: HEMMED ☐ CURLED ☐ EMBOSSED ☒ BEADED BORDER ☐ STAMP FRAME ☒ RIGHT ANGLE BEND ☐ BACK FRAME ☐ PACKING: PER CARTON <u>10</u> PER CRATE <u>10</u> PER BUNDLE <u>10</u> OTHER: <u>Price is based on reproduction of customer supplied "Pack" box.</u> <u>Tooling: Form die, brass emboss die to achieve detail on eagle.</u>					
<table style="width: 100%;"> <tr> <td>QUANTITIES: <u>500 Plaques</u></td> <td>One time tooling @ \$3,015.00</td> </tr> <tr> <td>PRICE: <u>\$9.18 each</u></td> <td>Steel tips <u>\$1,045.00</u></td> </tr> </table>		QUANTITIES: <u>500 Plaques</u>	One time tooling @ \$3,015.00	PRICE: <u>\$9.18 each</u>	Steel tips <u>\$1,045.00</u>
QUANTITIES: <u>500 Plaques</u>	One time tooling @ \$3,015.00				
PRICE: <u>\$9.18 each</u>	Steel tips <u>\$1,045.00</u>				
BILLING: ☒ BILL AS MANUFACTURE. ☐ BILL AS SHIP FOR. ☐ 6 MOS. ☐ 12 MOS. WAREHOUSING: ☒ SHIP IMMEDIATELY. ☐ 6 MOS. WHSE. ☐ 12 MOS. WHSE. DROP SHIPPING: <u> </u> PER SHIPMENT.					
CONDITIONS: PRICES QUOTED HEREIN ARE BASED ON CURRENT COSTS AND ARE FOR IMMEDIATE ACCEPTANCE. *These prices will be updated periodically, from date of order, to reflect changes in material and labor costs. Prices and terms quoted herein are not subject to verbal charges unless confirmed in writing by the home office of this company in St. Louis, MO. All contracts for delivery are contingent upon fees, strikes or other causes beyond our control. Purchaser agrees to accept increases or reductions in the amount of 10% of the quantity ordered. Any taxes or governmental charges which the seller may be required to pay or collect upon the production, sale, storage, or delivery, under any existing or future law shall be for the account of the buyer who shall promptly pay the amount billed to the seller upon demand. F.O.B. ST. LOUIS, MISSOURI. Shipments via truck-shipped freight collect. A 1% service charge will be added to freight costs for all prepaid shipments involving 5 or more pieces.					
<table style="width: 100%;"> <tr> <td style="width: 50%;"> TERMS — NET 10 DAYS A service charge of 1 1/2% per month will be applied to all unpaid balances over 30 days. </td> <td style="width: 50%; text-align: center;"> STOUT INDUSTRIES, INC. </td> </tr> </table>		TERMS — NET 10 DAYS A service charge of 1 1/2% per month will be applied to all unpaid balances over 30 days.	STOUT INDUSTRIES, INC. 		
TERMS — NET 10 DAYS A service charge of 1 1/2% per month will be applied to all unpaid balances over 30 days.	STOUT INDUSTRIES, INC. 				

87526380

Figura 6: Exemplo de página do dataset FUNSD [15]

Como a quantidade de imagens nas duas bases é muito pequena, é feito uma etapa de *data augmentation* para aumentar o número de imagens e prover imagens de diversos níveis de qualidade durante a etapa de treinamento. Para treinar as redes neurais utilizadas na etapa de avaliação [37][29][27] de qualidade é utilizado uma divisão da base de dados em 80% dos dados para treinamento e 20% dos dados para teste na base de dados StaVer e

75% dos dados para treinamento e 25% dos dados para teste na base de dados FUNSD. Os métodos que não utilizam treinamento [25] são executados somente nos dados de teste, para ser possível uma comparação de igual para igual com os métodos de aprendizado de máquina.

É necessário notar também que cada imagem anotada tem três componentes, base, tabelas e carimbos. O treinamento de cada modelo e as inferências de cada teste no método proposto são feitas por componente, fazendo com que cada *dataset* seja virtualmente composto de três sub-divisões, do tipo *StaVer-Tables*, *StaVer-Stamps* e *StaVer-Base*. Na fase de treinamento das redes utilizadas será necessário separar os modelos por cada componente onde o mesmo for treinado, fazendo com que para cada rede tenha um modelo treinado para tabelas, um modelo treinado para carimbos e um modelo treinado para documento base. Vale ressaltar que a base de dados FUNSD não tem carimbos, então as redes treinadas na mesma só terão dois modelos, um para tabelas e outro para documento base.

Foram decididos manter somente essas duas bases de dados durante a experimentação, pois as mesmas são as mais alinhadas com a definição de documento técnico dado no Capítulo 1 e que também contêm algum tipo de degradação nas suas imagens, seja por ajuste na qualidade do *scanner* ou por degradação ao imprimir a página. Como o intuito da proposta é medir a qualidade de uma página é necessário a variação de qualidade das mesmas, para ser possível identificar se o método proposto lida bem com imagens de qualidades variadas e não somente com imagens de uma qualidade só. Bases de dados como SROIE[13], PubTables-1M[42] e PubLayNet[47] também não foram levantadas ou por não conter nenhum dos componentes de tabela ou carimbo, no caso do SROIE, ou por não ter nenhuma variedade na qualidade, no caso do PubTables-1M e PubLayNet.

3.2 Data Augmentation

Para a utilização de *data augmentation* nas bases de dados é necessário focar em degradações que são provadas a afetar sistemas OCR [7, 14, 5]. Sendo assim foram utilizados 6 tipos de degradações diferentes contidas dentro da biblioteca Augraphy [12], que não só mostra ter efeito em processos OCR, mas também foca em degradações que podem ocorrer em documentos que são escaneados e impressos, onde normalmente os processos de automatização de tarefas utilizando OCR ocorrem. Todas os exemplos de imagens degradadas nas próximas subseções terão base na Figura 7. As próximas subseções caracterizam cada variação de degradação utilizadas durante o projeto, com 12 variações no total criando uma base de dados ampliada com 12.480 imagens para o StaVer e 650 para o FUNSD, valendo ressaltar que as técnicas de *data augmentation* a seguir foram utilizadas somente nos dados de treinamento para melhorar a generalidade dos regressores treinados no trabalho.

Martin Schreiber, Bergstraße 242, 06289 Mühltröff

SarahSara Maier
Poststraße 170
56464 Stadt Wehlen

Ihre Bestellung vom
1. Juli 2005

LIQUIDATION

Sehr geehrte Frau Maier,

Bitte überweisen Sie den folgenden Rechnungsbeitrag innerhalb von 7 Tagen auf das unten angegebene Konto.

Anzahl	Beschreibung	Einzelpreis	Gesamtpreis
1	Versandkosten	11.03	11,03 €
9	50m Kabeltrommel, Überspannungsschutz	25.99	233,91 €
12	Mike-Er Turnschuhe, Gr.45	109.99	1319,88 €
4	Radiowecker	13.86	55,44 €
8	Tiefkühlpizza 10er Packung	19.99	159,92 € ⁺
18	Porzellanteller, klein	2.23	40,14 €
6	Orangenlimonade	2.06	12,36 €
12	Koffeinhaltiges Getränk, 1,5 l	2.16	25,92 €
29	Riesling, Mosel, 0,75 l	2.45	71,05 €
20	Panflöte	101.99	2039,80 €
5	Drahtlos Tastatur und Maus, Set DS13d	56.78	283,90 €
29	Vollkornbrot 0,5 kg	1.11	32,19 € ⁺
18	Stereokopfhörer ST-22	32.56	586,08 €
25	Router 16Port UpLinx	123.98	3099,50 €
9	Katalog 2010	0.50	4,50 €
15	Taschenrechner	23.99	359,85 €
23	Tee, Early Grey	1.99	45,77 €

Mensagem na Garrafa
Martin Schreiber

Bergstraße 242
06289 Mühltröff

Telephon: 0670/2272
Fax: 0670/2272

Ust-Id: DE 1307511670

Datum
15. November 2010

Gesellschafter:
Martin Schreiber
UStId-Nr
DE 1307511670

Geschäftsführer:
Martin Schreiber
Gerichtsstand:
Hohenleuben

Figura 7: Exemplo de página sem utilização de *data augmentation*.

3.2.1 *Inkbleed*

Usa uma detecção de arestas Sobel para criar uma máscara de todas as arestas e, em seguida, aplica ruído a essas arestas. A Figura 8 demonstra o efeito da degradação. Neste projeto foi variado esta degradação com três tamanhos de *kernel*, sendo os mesmos um de tamanho (3,3), outro de tamanho (5,5) e o último de tamanho (7,7).

Mensagem na Garrafa
Martin Schreiber

Bergstraße 242
06289 Mühltröff

Telephon: 0670/2272
Fax: 0670/2272

Ust-Id: DE 1307511670

Martin Schreiber, Bergstraße 242, 06289 Mühltröff

SarahSara Maier
Poststraße 170
56464 Stadt Wehlen

Ihre Bestellung vom
1. Juli 2005

Datum
15. November 2010

LIQUIDATION

Sehr geehrte Frau Maier,

Bitte überweisen Sie den folgenden Rechnungsbeitrag innerhalb von 7 Tagen auf das unten angegebene Konto.

Anzahl	Beschreibung	Einzelpreis	Gesamtpreis
1	Versandkosten	11.03	11,03 €
9	50m Kabeltrommel, Überspannungsschutz	25.99	233,91 €
12	Mike-Er Turnschuhe, Gr.45	109.99	1319,88 €
4	Radiowecker	13.86	55,44 €
8	Tiefkühlpizza 10er Packung	19.99	159,92 € ¹
18	Porzellanteller, klein	2.23	40,14 €
6	Orangenlimonade	2.06	12,36 €
12	Koffeinhaltiges Getränk, 1,5 l	2.16	25,92 €
29	Riesling, Mosel, 0,75 l	2.45	71,05 €
20	Pantlöte	101.99	2039,80 €
5	Drahtlos Tastatur und Maus, Set DS13d	56.78	283,90 €
29	Vollkornbrot 0,5 kg	1.11	32,19 € ²
18	Stereokopfhörer ST-22	32.56	586,08 €
25	Router 16Port UpLinx	123.98	3099,50 €
9	Katalog 2010	0.50	4,50 €
15	Taschenrechner	23.99	359,85 €
23	Tee, Early Grey	1.99	45,77 €

Bankverbindung:
Konto 667 379 181 138
Deutsche Bank
BLZ 643 292 08

Gesellschafter:
Martin Schreiber
UStId-Nr
DE 1307511670

Geschäftsführer:
Martin Schreiber
Gerichtsstand:
Hohenleuben

Figura 8: Exemplo de página utilizando *Inkbleed* com tamanho de *kernel* (5,5).

3.2.2 *Folding*

Simula um efeito de dobramento que pode ocorrer durante mal posicionamento de um papel ao escanear ou imprimir um documento. A Figura 9 demonstra o efeito da degradação. Neste projeto esta degradação não foi variada, deixando a mesma com um padrão de 5 dobramentos na página para todas as páginas da base de dados.

Martin Schreiber, Bergstraße 242, 06289 Mühltröff

SarahSara Maier
Poststraße 170
56464 Stadt Wehlen

Ihre Bestellung vom
1. Juli 2005

LIQUIDATION

Sehr geehrte Frau Maier,

Bitte überweisen Sie den folgenden Rechnungsbeitrag innerhalb von 7 Tagen auf das unten angegebene Konto.

Anzahl	Beschreibung	Einzelpreis	Gesamtpreis
1	Versandkosten	11.03	11.03 €
9	50m Kabeltrommel, Überspannungsschutz	25.99	233.91 €
12	Mike-Er Turnschuhe, Gr.45	109.99	1319.88 €
4	Radiowecker	13.86	55.44 €
8	Tiefkühlpizza 10er Packung	19.99	159.92 €
18	Porzellanteller, klein	2.23	40.14 €
6	Orangenlimonade	2.06	12.36 €
12	Koffeinhaltiges Getränk, 1,5 l	2.16	25.92 €
29	Riesling, Mosel, 0,75 l	2.45	71.05 €
20	Panflöte	101.99	2039.80 €
5	Drahtlos Tastatur und Maus, Set DS13d	56.78	283.90 €
29	Vollkornbrot 0,5 kg	1.11	32.19 €
18	Stereokopfhörer ST-22	32.56	586.08 €
25	Router 16Port UpLinx	123.98	3099.50 €
9	Katalog 2010	0.50	4.50 €
15	Taschenrechner	23.99	359.85 €
23	Tee, Early Grey	1.99	45.77 €

**Messagen
na Garrafa**
Martin Schreiber

Bergstraße 242
06289 Mühltröff

Telephon: 0670/2272
Fax: 0670/2272

Ust-Id: DE 1307511670

Datum
15. November 2010

Bankverbindung:
Konto 687 379 181 138
Deutsche Bank
BLZ 643 292 68

Gesellschafter:
Martin Schreiber
UStId-Nr
DE 1307511670

Geschäftsführer:
Martin Schreiber
Gerichtsstand:
Hohenleuben

Figura 9: Exemplo de página utilizando *Folding*.

3.2.3 DirtyDrum

Simula de sujeira nos tambores de uma impressora a toner, criando faixas de ruído na página impressa. As Figuras 10, 11 e 12 demonstram o efeito da degradação. Neste projeto esta degradação tem três variações, incluindo imagens com faixas de ruído horizontal, imagens com faixas de ruído verticais e imagens com as duas orientações de faixa de ruído ao mesmo tempo.



Figura 10: Exemplo de página utilizando *DirtyDrum* de forma vertical.

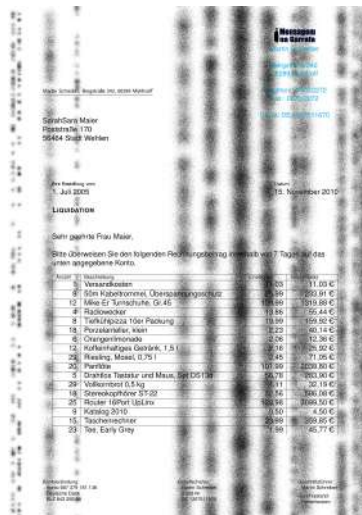


Figura 11: Exemplo de página utilizando *DirtyDrum* de forma horizontal.

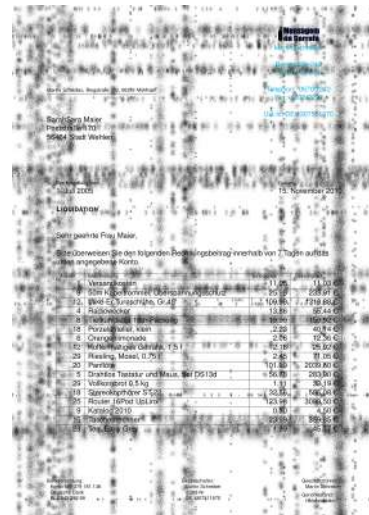


Figura 12: Exemplo de página utilizando *DirtyDrum* de forma vertical e horizontal.

3.2.4 Letterpress

Simula um efeito de tinta impressa em quantidade desigual, fazendo com que a cor em partes do texto fique mais fraca em relação a outras áreas do texto. A Figura 13 demonstra o efeito da degradação. Neste projeto esta degradação não foi variada, deixando a mesma com os valores padrão da biblioteca Augraphy.

3.2.5 Artefatos JPEG

Cria artefatos de compressão na imagem deixados ao utilizar o formato JPEG com diferente níveis de qualidade. As Figuras 14, 15 e 16 demonstram o efeito da degradação. Neste projeto esta degradação tem três variações com base na qualidade da compressão JPEG, sendo a primeira com 50% de qualidade, a segunda com 10% e a última com 1% de qualidade

3.2.6 Scribbles

Aplica um efeito de rabiscos sobre a página, simulando problemas como assinaturas e outros tipos de ruídos causados fora do processo de escaneamento/impressão. A Figura

Martin Schreiber, Bergstraße 242, 06289 Mühltröff

SarahSara Maier
Poststraße 170
56464 Stadt Wehlen

Ihre Bestellung vom
1. Juli 2005

LIQUIDATION

Sehr geehrte Frau Maier,

Bitte überweisen Sie den folgenden Rechnungsbeitrag innerhalb von 7 Tagen auf das unten angegebene Konto.

**Mensagem
na Garrafa**
Martin Schreiber

Bergstraße 242
06289 Mühltröff

Telephon: 0670/2272
Fax: 0670/2272

Ust-Id: DE 1307511670

Datum
15. November 2010

Anzahl	Beschreibung	Einzelpreis	Gesamtpreis
1	Versandkosten	11.03	11.03 €
9	50m Kabeltrommel, Überspannungsschutz	25.99	233.91 €
12	Mike-Er Turnschuhe, Gr.45	109.99	1319.88 €
4	Radiowecker	13.86	55.44 €
8	Tiefkühlpizza 10er Packung	19.99	159.92 €
18	Porzellanteller, klein	2.23	40.14 €
6	Orangenlimonade	2.06	12.36 €
12	Koffeinhaltiges Getränk, 1,5 l	2.16	25.92 €
29	Riesling, Mosel, 0,75 l	2.45	71.05 €
20	Panflöte	101.99	2039.80 €
5	Drahtlos Tastatur und Maus, Set DS13d	56.78	283.90 €
29	Vollkornbrot 0,5 kg	1.11	32.19 €
18	Stereokopfhörer ST-22	32.56	586.08 €
25	Router 16Port UpLinx	123.98	3099.50 €
9	Katalog 2010	0.50	4.50 €
15	Taschenrechner	23.99	359.85 €
23	Tee, Early Grey	1.99	45.77 €

Bankverbindung:
Konto 687 379 101 138
Deutsche Bank
BLZ 643 292 68

Gesellschafter:
Martin Schreiber
UStId-Nr
DE 1307511670

Geschäftsführer:
Martin Schreiber
Gerichtsstand:
Hohenleuben

Figura 13: Exemplo de página utilizando *Letterpress*.

The figure shows three versions of a document page, each representing a different level of JPEG compression quality (50%, 10%, and 1%). The document is a 'Liquation' form from 'Messe der Stadt'. The text is mostly legible, but the images and some text elements show increasing degradation as the compression quality decreases. The tables at the bottom of each page contain numerical data, which remains mostly readable even at 1% quality.

Figura 14: Exemplo de página utilizando compressão JPEG com 50% de qualidade.

Figura 15: Exemplo de página utilizando compressão JPEG com 10% de qualidade.

Figura 16: Exemplo de página utilizando compressão JPEG com 1% de qualidade.

17 demonstra o efeito da degradação. Neste projeto esta degradação não foi variada, deixando a mesma com os valores padrão da biblioteca Augraphy.

3.3 Motores OCR

Todos os métodos de DIQA objetivos sem referência necessitam de motores OCR para averiguar a correlação entre a métrica de avaliação predita e a acurácia real do OCR. Neste trabalho são utilizados dois motores de OCR, Tesseract[41] e PaddleOCR[9]. As próximas duas subseções entram em detalhes sobre os dois motores selecionados

3.3.1 Tesseract

Tesseract[41] é um *software* de OCR de código-aberto que conta com modelos treinados para diversas linguagens. Este sistema foi escolhido pela sua capacidade de reconhecer múltiplas linguagens, o que é necessário devido as bases de dados do projeto conterem mais de uma linguagem, e oferecer customização de seus parâmetros e retreinamento de seus modelos caso necessário. As configurações utilizadas no projeto são:

- Versão: 5.2.0
- Método de *thresholding*: Sauvola
- DPI: 300
- Modelo de linguagem: Latin
- PSM: 11 (Ignorar a análise de página interna do Tesseract)

Martin Schreiber, Bergstraße 242, 06289 Mühltröff

SarahSara Maier
Poststraße 170
56464 Stadt Wehlen

Ihre Bestellung vom
1. Juli 2005

LIQUIDATION

Sehr geehrte Frau Maier,

Bitte überweisen Sie den folgenden Rechnungsbeitrag innerhalb von 7 Tagen auf das unten angegebene Konto.

Anzahl	Beschreibung	Einzelpreis	Gesamtpreis
1	Versandkosten	11.03	11,03 €
9	50m Kabeltrommel, Überspannungsschutz	25.99	233,91 €
12	Mike-Er Turnschuhe, Gr.45	109.99	1319,88 €
4	Radiowecker	13.86	55,44 €
8	Tiefkühlpizza 10er Packung	19.99	159,92 € ⁺
18	Porzellanteller, klein	2.23	40,14 €
6	Orangenlimonade	2.06	12,36 €
12	Koffeinhaltiges Getränk, 1,5 l	2.16	25,92 €
29	Riesling, Mosel, 0,75 l	2.45	71,05 €
20	Panflöte	101.99	2039,80 €
5	Drahtlos Tastatur und Maus, Set DS13d	56.78	283,90 €
29	Vollkornbrot 0,5 kg	1.11	32,19 € ⁺
18	Stereokopfhörer ST-22	32.56	586,08 €
25	Router 16Port UpLinx	123.98	3099,50 €
9	Katalog 2010	0.50	4,50 €
15	Taschenrechner	23.99	359,85 €
23	Tee, Early Grey	1.99	45,77 €

Mensagem na Garrafa
Martin Schreiber

Bergstraße 242
06289 Mühltröff

Telephon: 0670/2272
Fax: 0670/2272

Ust-Id: DE 1307511670

Datum
15. November 2010

Bankverbindung:
Konto 687 379 181 138
Deutsche Bank
BLZ 643 292 68

Gesellschafter:
Martin Schreiber
UStId-Nr
DE 1307511670

Geschäftsführer:
Martin Schreiber
Gerichtsstand:
Hohenleuben

Figura 17: Exemplo de página utilizando *Scribbles*.

- OEM: 1 (Rede LSTM)

Antes de utilizar este OCR foi feito um estudo com base em artigos existentes que marcavam a performance do Tesseract pior do que outros métodos do estado da arte[34], onde no mesmo foi feito um teste utilizando a métrica CLEval[2] na base de dados FUNSD[15] com a configuração do Tesseract acima e foram encontrados os resultados da tabela 1, que mostram o Tesseract com resultados não tão distantes a outros métodos do estado da arte se o mesmo for configurado de maneira correta.

Tabela 1: Resultados de acurácia do TesseractOCR - FUNSD

	Precisão	Recall	F1
Detecção	95,07%	87,13%	90,93%
End-to-End	88,54%	90,97%	89,74%

3.3.2 Paddle

PaddleOCR[9] é um sistema de OCR de código aberto desenvolvido pela equipe PaddlePaddle. Este sistema foi escolhido tanto pela sua capacidade de reconhecer múltiplas linguagens quanto pela sua melhor performance em documentos técnicos comparado ao Tesseract, o que pode mostrar diferenças no método de avaliação de qualidade final dependendo da qualidade do motor de OCR utilizado. As configurações utilizadas no projeto são:

- Versão: 2
- Rede de detecção de texto: *Differentiable Binarization* (DBNet[28])
- : Rede de reconhecimento de texto: *Convolutional Recurrent Neural Network* (CRNN[23])
- Modelo de linguagem/reconhecimento: Inglês para FUNSD e Alemão para StaVer

O mesmo teste feito na base de dados FUNSD foi estendido para o PaddleOCR, com os resultados na tabela 2. É possível notar que o mesmo é consideravelmente melhor que o Tesseract tanto na detecção quanto no reconhecimento de caracteres, o que confirma a hipótese inicial de variação de qualidade de motor OCR.

Tabela 2: Resultados de acurácia do PaddleOCR - FUNSD

	Precisão	Recall	F1
Detecção	99,15%	98,53%	98,84%
End-to-End	98,73%	97,42%	98,07%

3.4 Métricas de qualidade de OCR

Antes de utilizar os *datasets* é necessário primeiro calcular uma métrica de qualidade para cada componente de cada imagem do mesmo. Esta etapa é feita após ser aplicado *data augmentation* em cada imagem, sendo que as imagens são separadas em seus três elementos (base, carimbo e tabela) e, então, processadas por um sistema OCR. A saída do sistema OCR é comparada com o *ground-truth* para aquela imagem, gerando uma métrica de acurácia para o processo OCR naquela imagem, onde essa métrica de acurácia também é tomada como a métrica de qualidade da imagem para que seja possível utilizar as técnicas de DIQA na mesmas. Para parear o texto de saída do OCR com o texto do *ground-truth* é feito uma processo de comparação entre cada linha encontrada na saída OCR e cada linha no texto anotado, calculando a intersecção-sobre-união (IoU) e pareando as duas linhas se a mesma for maior que 0.5.

As métricas a serem exploradas nesse projeto são:

- Taxa de erro de caracteres: *Character Error Rate* (CER) utiliza uma métrica de distância, no caso desse trabalho a distância de Levenshtein [24], para medir a quantidade de manipulações (substituições, inserções e remoções) necessárias para transformar um texto em outro texto. O CER faz a soma dessa distância de edição entre as linhas pareadas e divide o valor pelo número de caracteres na linha para normalizar o mesmo. Esta métrica foca em quantificar o quão bom o reconhecimento de texto de um sistema OCR é.
- CLEval [2]: é uma métrica de desempenho de OCR que utiliza de "sub-métricas" para medir a performance de um sistema OCR, onde essas "sub-métricas" contêm técnicas para análise de detecção de OCR, medindo *overlaps* e *merges* nas caixas de texto encontradas pelo OCR, assim como técnicas para a análise do reconhecimento de OCR, medindo características como o número de caracteres corretamente encontrados. Esta métrica *end-to-end* foca no processo inteiro de OCR, desde a detecção de texto até o reconhecimento do mesmo.

3.5 Métodos de DIQA

Os métodos de DIQA escolhidos serão utilizados nos três componentes de cada imagem, sendo os mesmos selecionados pela sua acurácia na tarefa de DIQA em comparação com o resto do estado da arte e na sua adaptabilidade a novos problemas. Quatro métodos de DIQA serão utilizados para a pesquisa, valendo ressaltar que os três primeiros foram explicados durante a seção de trabalhos relacionados:

- Transfer-DIQA [29]: método de DIQA utilizando transferência de aprendizado de uma rede CNN. Essa rede está sendo utilizada para representar o estado da arte em CNN's utilizadas para a tarefa de DIQA.
- Attention-DIQA [27]: método de DIQA utilizando uma rede CNN. Essa rede está sendo utilizada para representar o estado da arte em RNN's utilizadas para a tarefa de DIQA.
- CG-DIQA [25]: método de DIQA utilizando técnicas clássicas de visão computacional. Essa técnica está sendo utilizada no projeto para representar o estado da arte em técnicas clássicas de DIQA. Vale a pena ressaltar que nas publicações de DIQA o método se mantém competitivo embora o mesmo não utilize aprendizado de máquina.
- Transfer-Efficient-DIQA: método de DIQA baseado no Transfer-DIQA, porém no lugar da AlexNet utilizada no mesmo é feito uma substituição pela EfficientNet-B0, que é uma rede mais atual, robusta e precisa. O *fine-tuning* da mesma é igual ao da AlexNet, com um modelo pré-treinado na ImageNet-1k transformado para

regressão ao tirar o último bloco *fully connected* (FC) de 1000 neurônios e substituir o mesmo com um bloco *fully connected* de um neurônio e um regressor baseado em perda L2, e retreinado a rede com base na tarefa de DIQA. Após o primeiro *fine-tuning* são congeladas as camadas convolucionais iniciais e adicionados três camadas diretamente antes da camada final de regressão: uma FC de 5 neurônios seguida por duas camadas FC de 10 neurônios. Essas camadas são treinadas do zero para aprender características específicas a tarefa de DIQA. Assim como no Transfer-DIQA, as imagens de entrada são redimensionadas para 224 de altura por 224 de largura em pixels para o processamento na rede.

Vale ressaltar que redes Transformer não estão sendo consideradas porque as mesmas ainda não se mostraram com resultados competitivos em relação ao estado da arte na área de DIQA.

O método CG-DIQA [25] não é um método de aprendizado de máquina, então o mesmo não precisa de treinamento. Os outros dois métodos necessitam de treinamento, onde o processo de treinamento é descrito abaixo.

O fluxograma de treinamento das redes de DIQA pode ser visto na figura 18. A qualidade que a rede tenta fazer a previsão é dada por uma métrica que vem da comparação entre a saída do OCR e o texto original da imagem, onde a seleção desta métrica é discutida com mais detalhes na seção 3.6. Após o treinamento a rede tenta fazer a regressão deste valor para uma imagem nova. O fluxograma de execução da rede pode ser visto na figura 19. No final de tudo a entrada da rede é uma imagem e a saída é uma previsão do valor da métrica utilizada, que no caso atual é um valor entre 0 e 1. Vale também ressaltar que esse treinamento é feito por componente, isto é, separando tabelas, carimbos e base do documento e treinando um modelo específico para cada análise de qualidade de cada um dos mesmos.

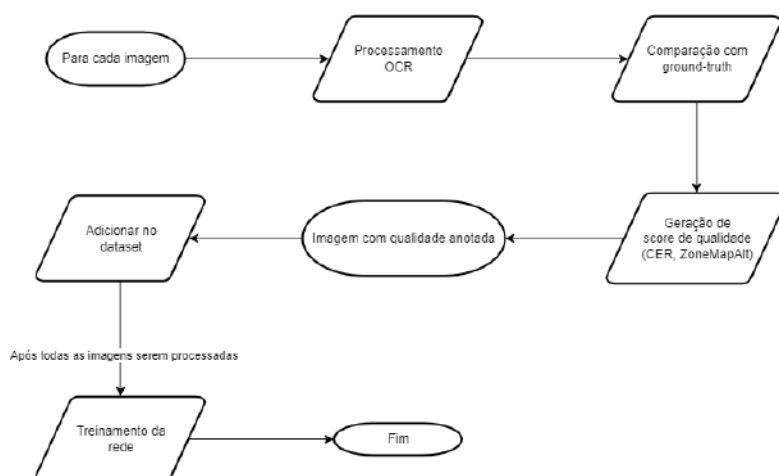


Figura 18: Treinamento de uma rede DIQA.

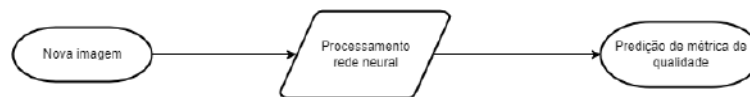


Figura 19: Execução de uma rede DIQA.

3.6 Métrica de saída do OCR

Como mencionado na seção 2 os trabalhos de DIQA atuais focam no impacto da qualidade da imagem no reconhecimento do texto em uma página sem olhar no impacto que o mesmo tem na detecção deste texto, normalmente utilizando métricas como a taxa de erro de caracteres, ignorando a primeira parte de sistemas de OCR. Como a maioria dos processos de visão computacional a detecção de texto também sofre com perdas devido da degradação da imagem a ser processada, portanto para tentar adicionar essa relação de qualidade e detecção de texto foi utilizado a métrica de *Recognition Score End-to-End* [2], uma métrica que combina o resultado da detecção e do reconhecimento do OCR para analisar a qualidade da saída do mesmo.

A métrica de saída é importante em dois pontos:

1. Para servir como variável de regressão durante o treinamento de métodos com aprendizado de máquina.
2. Para analisar a correlação entre os valores preditos e os valores corretos do sistema.

Para conseguir fazer uma comparação válida entre todos os métodos é aplicado um processo de correlação entre o *Recognition Score* de cada documento de teste e seu valor de qualidade predito. Estes métodos de correlação fazem com que seja possível comparar a qualidade predita da qualidade real, assim como normalizando o dado entre 0 e 1, possibilitando assim a comparação justa entre diferentes métodos de DIQA. Para fazer essa correlação são elencados dois coeficientes de correlação distintos que serão explicados nas próximas duas subseções. Vale ressaltar que como coeficientes de correlação quanto maior o valor obtido melhor é o método de DIQA.

3.6.1 Linear Correlation Coefficient

Linear Correlation Coefficient (LCC) [11] ou *Pearson Correlation Coefficient* é um coeficiente de correlação linear entre dois conjuntos de dados, sendo utilizados neste trabalho o conjunto de dados “X” como a métrica de qualidade real dos documentos testados e o conjunto de dados “Y” como a métrica de qualidade predita para os documentos. A equação 1 demonstra a fórmula do coeficiente linear, onde:

- x_i e y_i são valores individuais dos conjuntos X e Y, respectivamente;
- n é o tamanho da amostra de dados;

- $\bar{x}$ e $\bar{y}$ são a média de valores dos conjuntos X e Y, respectivamente;
- s_x e s_y são os desvios padrões dos conjuntos X e Y, respectivamente.

$$r_{xy} = \frac{1}{n-1} \sum_{i=1}^n \left(\frac{x_i - \bar{x}}{s_x} \right) \left(\frac{y_i - \bar{y}}{s_y} \right) \quad (1)$$

3.6.2 Spearman Rank Order Correlation Coefficient

Spearman Rank Order Correlation Coefficient (SROCC) [8] é uma medida de correlação de classificação que avalia o quão bem a relação entre dois conjuntos de dados pode ser descrita usando uma função monotônica. Neste trabalho o conjunto de dados “X” como a métrica de qualidade real dos documentos testados e o conjunto de dados “Y” como a métrica de qualidade predita para os documentos. A equação 2 demonstra a fórmula do coeficiente de Spearman, onde:

- $R(X)$ e $R(Y)$ são os valores de X e Y convertidos para *ranks*, respectivamente;
- $cov(R(X) R(Y))$ é a covariância entre os *ranks*;
- $\sigma_{R(X)}$ e $\sigma_{R(Y)}$ são os desvios padrões das variáveis de *rank* X e Y, respectivamente.

$$r_s = \frac{cov(R(X) R(Y))}{\sigma_{R(X)} \sigma_{R(Y)}} \quad (2)$$

3.7 Fusão de valores de qualidade

Ao final do processo haverá diversos valores de qualidade para cada componente analisado durante o mesmo. Para atribuir um valor final ao mesmo é necessário agregar estes valores através de uma função. O processo atual se divide em duas partes:

1. Agregar os valores de componentes do mesmo tipo. Exemplo: duas tabelas em uma mesma página.
2. Agregar os valores dos três tipos de componentes (tabela, carimbo e base).

Diversos métodos para a combinação de resultados de regressores existem na literatura [1], porém nesse trabalho foram adotadas seis funções de fusão, sendo elas:

- Média: função básica de fusão para servir como linha de base para a análise, adiciona todos os valores de entrada e divide pela quantidade dos mesmos.
- Média ponderada: média com pesos, onde os pesos para cada valor se dão como a intersecção sobre a união da área do componente em relação a área total do documento.
- Máximo: atribui o valor máximo entre os valores de entrada como valor correto.

- Mínimo: atribui o valor mínimo entre os valores de entrada como valor correto.
- Mediana: encontra a mediana entre os valores de entrada e coloca ele como saída.
- Média geométrica: calcula o produto de todos os valores de entrada elevado ao inverso da quantidade dos mesmos.

Estas funções foram adotadas por sua eficácia na literatura quando o problema a ser resolvido requer um número baixo de valores a serem agregados, sendo que no caso deste projeto a quantidade de valores a serem agregados tende a girar em torno de três, devido ao comportamento de dividir e conquistar em três componentes distintos.

Ao final o método a ser aplicado terá as etapas:

1. Segmentar imagem em tabelas, carimbos e base;
2. Aplicar método de DIQA em cada segmento;
3. Fazer a fusão dos valores de qualidade entre componentes do mesmo tipo;
4. Fazer a fusão dos valores de qualidade entre tabela, carimbo e base.

No final disso o processamento proposto terá gerado um valor de qualidade geral para um documento.

4 RESULTADOS

Os resultados do processo são dados para cada par de motor OCR e base de dados, onde a parte de teste da base de dados tem cada imagem analisada pelas 4 técnicas de DIQA descritas na seção de metodologia, com os seus modelos treinados nos dados de treinamento da mesma base, sem o processo de divisão-e-conquista por componentes da página para obter um valor base do estado da arte atual naquele *dataset*. Após isso é aplicado o método de divisão-e-conquista discutido na metodologia, onde cada imagem de teste é decomposta em seus três componentes e as técnicas de DIQA são aplicados nos mesmos para depois ser feita a fusão dos valores encontrados pelas seis funções (média, média ponderada, máximo, mínimo, mediana e média geométrica) para obter um valor de qualidade predito da imagem original. No fim todos os valores de qualidade preditos são comparados com os valores de acurácia de OCR para medir a correlação entre os mesmos, mostrando também o grau de acurácia daquele método DIQA.

As próximas seções detalharam e discutiram os resultados para cada par motor OCR e base de dados, mostrando alguns experimentos complementares de teste de modelos *cross-dataset* e do impacto que as *data augmentations* tem no modelo e, por fim, demonstrando um exemplo qualitativo do processo de divisão-e-conquista em uma página do *dataset* StaVer.

4.1 StaVer

Os resultados da base de dados StaVer se dão em nas próximas subseções para cada motor de OCR. Vale ressaltar que, diferente da base de dados FUNSD, as imagens da base de dados tem todos os componentes anotados, contando com tabelas, carimbo e base, fazendo com que a fusão de métricas seja dada utilizando todas as três métricas.

4.1.1 Tesseract

Antes de analisarmos os resultados totais é necessário detalhar quais das 4 técnicas DIQA são melhores em quais componentes, para no final conseguirmos utilizar as melhores técnicas para cada componente durante a fusão de métricas.

Os resultados da tabela 3 demonstram a correlação linear (LCC) de cada componente com as técnicas DIQA, com a tabela 4 fazendo o mesmo para a correlação monotônica (SROCC). Pode ser visto nesses resultados que a técnica Transfer-Efficient-DIQA tende a ser o melhor candidato para todos os componentes, com Transfer-DIQA normalmente em segundo lugar. Uma exceção a isso é no componente "Base", onde a técnica CG-DIQA tende a ter uma acurácia bem elevada, provavelmente ocasionado porque a métrica CG-DIQA foi feita para ser utilizada em documentos sem essas formas complexas. Outra curiosidade que podemos notar é a qualidade de tabelas, que normalmente é um ponto fraco do Tesseract, onde a técnica sem aprendizado CG-DIQA e a técnica de RNN se saem mal, provavelmente por que embora a imagem da tabela tenha qualidade boa o motor de OCR em si não tem se comporta bem no mesmo, onde provavelmente o *fine-tuning* e criação de camadas nas técnicas de *transfer learning* ajudam ao aprendizado dessa peculiaridade.

Tabela 3: Correlação de técnicas por componente - Par Tesseract e StaVer - LCC

Técnica	Base	Tabela	Carimbo
Transfer-DIQA	0,78	<u>0,81</u>	<u>0,79</u>
Transfer-Efficient-DIQA	<u>0,84</u>	0,89	0,83
Attention-DIQA	0,72	0,67	0,76
CQ-DIQA	0,85	0,59	0,75

Tabela 4: Correlação de técnicas por componente - Par Tesseract e StaVer - SROCC

Técnica	Base	Tabela	Carimbo
Transfer-DIQA	0,75	<u>0,80</u>	<u>0,78</u>
Transfer-Efficient-DIQA	0,83	0,88	0,82
Attention-DIQA	0,69	0,66	0,74
CQ-DIQA	<u>0,81</u>	0,62	0,73

Com isso em mente a tabela 5 testará os métodos de fusão de métricas utilizando a técnica Transfer-Efficient-DIQA para todos os componentes, mostrando também a linha de base de cada técnica sem a fusão de métricas.

Os resultados obtidos mostram que existe um aumento significativo, perto de 14% para o melhor resultado com um pulo de 0,84 para 0,96 na correlação linear (LCC) e 0,87 para 0,93 na correlação monotônica (SROCC), em documentos técnicos ao utilizar a técnica de divisão e conquista de componentes nesse par de OCR e base de dados.

4.1.2 Paddle

Antes de analisarmos os resultados totais é necessário detalhar quais das 4 técnicas DIQA são melhores em quais componentes, para no final conseguirmos utilizar as melho-

Tabela 5: Resultados de qualidade DIQA - Par Tesseract e Staver

Técnica	LCC	SROCC
Transfer-DIQA sem fusão	0,79	0,82
Transfer-Efficient-DIQA sem fusão	0,84	0,87
Attention-DIQA sem fusão	0,70	0,66
CQ-DIQA sem fusão	0,61	0,60
Fusão de métricas		
Média	0,91	0,89
Média ponderada	0,96	0,93
Máximo	0,75	0,77
Mínimo	0,68	0,69
Mediana	0,90	0,88
Média geométrica	0,85	0,87

res técnicas para cada componente durante a fusão de métricas.

Os resultados da tabela 6 demonstram a correlação linear (LCC) de cada componente com as técnicas DIQA, com a tabela 7 fazendo o mesmo para a correlação monotônica (SROCC). Pode ser visto nesses resultados que a técnica Transfer-Efficient-DIQA tende a ser o melhor candidato para todos os componentes, com Transfer-DIQA normalmente em segundo lugar. O motor de OCR Paddle lida bem com situações de tabelas, o que causa uma melhora no modelo de Attention-DIQA comparado com os modelos de transferência de aprendizado, porém é possível notar uma piora significativa no método sem aprendizado, provavelmente causado por esperar um resultado pior do que o obtido por um OCR com a robustez do Paddle.

Tabela 6: Correlação de técnicas por componente - Par Paddle e StaVer - LCC

Técnica	Base	Tabela	Carimbo
Transfer-DIQA	0,76	0,79	0,76
Transfer-Efficient-DIQA	0,82	0,87	0,80
Attention-DIQA	0,70	0,74	0,61
CQ-DIQA	0,75	0,58	0,52

Tabela 7: Correlação de técnicas por componente - Par Paddle e StaVer - SROCC

Técnica	Base	Tabela	Carimbo
Transfer-DIQA	0,73	0,77	0,75
Transfer-Efficient-DIQA	0,81	0,85	0,79
Attention-DIQA	0,68	0,70	0,63
CQ-DIQA	0,74	0,53	0,55

Com isso em mente, a tabela 8 testará os métodos de fusão de métricas utilizando a

técnica Transfer-Efficient-DIQA para todos os componentes, mostrando também a linha de base de cada técnica sem a fusão de métricas.

Tabela 8: Resultados de qualidade DIQA - Par Paddle e Staver

Técnica	LCC	SROCC
Transfer-DIQA sem fusão	0,77	0,79
Transfer-Efficient-DIQA sem fusão	0,82	0,85
Attention-DIQA sem fusão	0,68	0,64
CQ-DIQA sem fusão	0,60	0,58
Fusão de métricas		
Média	0,89	0,87
Média ponderada	0,94	0,91
Máximo	0,73	0,75
Mínimo	0,66	0,67
Mediana	0,88	0,85
Média geométrica	0,82	0,84

Os resultados obtidos ainda mostram um aumento significativo, perto de 14% para o melhor resultado com um pulo de 0,82 para 0,94 na correlação linear (LCC) e 0,85 para 0,91 na correlação monotônica (SROCC), em documentos técnicos ao utilizar a técnica de divisão e conquista de componentes nesse par de OCR e base de dados.

4.2 FUNSD

Os resultados da base de dados FUNSD se dão nas próximas subseções para cada motor de OCR. Vale ressaltar que, diferente da base de dados StaVer, as imagens da base de dados têm menos componentes anotados, focando em tabelas e base, fazendo com que a fusão de métricas seja dada utilizando apenas essas duas métricas.

4.2.1 Tesseract

Antes de analisarmos os resultados totais, é necessário detalhar quais das 4 técnicas DIQA são melhores em quais componentes, para no final conseguirmos utilizar as melhores técnicas para cada componente durante a fusão de métricas.

Os resultados da tabela 9 demonstram a correlação linear (LCC) de cada componente com as técnicas DIQA, com a tabela 10 fazendo o mesmo para a correlação monotônica (SROCC). Pode ser visto nesses resultados que a técnica Transfer-Efficient-DIQA tende a ser o melhor candidato para todos os componentes, com Transfer-DIQA em segundo lugar. Diferente do *dataset* StaVer, a base de dados FUNSD tem diversas formas complexas além de tabelas e carimbos, como formulários e logos, o que parece piorar a acurácia da técnica CG-DIQA.

Com isso em mente, a tabela 11 testará os métodos de fusão de métricas utilizando a

Tabela 9: Correlação de técnicas por componente - Par Tesseract e FUNSD - LCC

Técnica	Base	Tabela
Transfer-DIQA	0,76	0,80
Transfer-Efficient-DIQA	0,81	0,88
Attention-DIQA	0,74	0,72
CG-DIQA	0,60	0,50

Tabela 10: Correlação de técnicas por componente - Par Tesseract e FUNSD - SROCC

Técnica	Base	Tabela
Transfer-DIQA	0,74	0,78
Transfer-Efficient-DIQA	0,80	0,86
Attention-DIQA	0,71	0,70
CG-DIQA	0,58	0,48

técnica Transfer-Efficient-DIQA para todos os componentes, mostrando também a linha de base de cada técnica sem a fusão de métricas.

Tabela 11: Resultados de qualidade DIQA - Par Tesseract e FUNSD

Técnica	LCC	SROCC
Transfer-DIQA sem fusão	0,77	0,79
Transfer-Efficient-DIQA sem fusão	0,81	0,86
Attention-DIQA sem fusão	0,71	0,70
CG-DIQA sem fusão	0,59	0,55
Fusão de métricas		
Média	0,87	0,85
Média ponderada	0,93	0,93
Máximo	0,72	0,73
Mínimo	0,64	0,65
Mediana	0,86	0,84
Média geométrica	0,80	0,82

Os resultados obtidos mostram um aumento significativo, com um crescimento de aproximadamente 14% para o melhor resultado, indo de 0,81 para 0,93 na correlação linear (LCC) e de 0,86 para 0,93 na correlação monotônica (SROCC), em documentos técnicos ao utilizar a técnica de divisão e conquista de componentes nesse par de OCR e base de dados.

4.2.2 Paddle

Antes de analisarmos os resultados totais, é necessário detalhar quais das 4 técnicas DIQA são melhores em quais componentes, para no final conseguirmos utilizar as melhores técnicas para cada componente durante a fusão de métricas.

Os resultados da tabela 12 demonstram a correlação linear (LCC) de cada componente com as técnicas DIQA, com a tabela 13 fazendo o mesmo para a correlação monotônica (SROCC). Pode ser visto nesses resultados que a técnica Transfer-Efficient-DIQA tende a ser o melhor candidato para todos os componentes, com Transfer-DIQA normalmente em segundo lugar. O motor de OCR Paddle lida bem com situações de tabelas, o que causa uma melhora no modelo de Attention-DIQA comparado com os modelos de transferência de aprendizado, mas o método CG-DIQA tem um desempenho notavelmente pior, evidenciando a ineficácia desse modelo em estruturas complexas de tabelas.

Tabela 12: Correlação de técnicas por componente - Par Paddle e FUNSD - LCC

Técnica	Base	Tabela
Transfer-DIQA	0,74	0,78
Transfer-Efficient-DIQA	0,80	0,86
Attention-DIQA	0,72	0,71
CG-DIQA	0,56	0,47

Tabela 13: Correlação de técnicas por componente - Par Paddle e FUNSD - SROCC

Técnica	Base	Tabela
Transfer-DIQA	0,71	0,76
Transfer-Efficient-DIQA	0,79	0,83
Attention-DIQA	0,69	0,69
CG-DIQA	0,55	0,45

Com isso em mente, a tabela 14 testará os métodos de fusão de métricas utilizando a técnica Transfer-Efficient-DIQA para todos os componentes, mostrando também a linha de base de cada técnica sem a fusão de métricas.

Tabela 14: Resultados de qualidade DIQA - Par Paddle e FUNSD

Técnica	LCC	SROCC
Transfer-DIQA sem fusão	0,75	0,78
Transfer-Efficient-DIQA sem fusão	0,80	0,79
Attention-DIQA sem fusão	0,71	0,69
CG-DIQA sem fusão	0,57	0,53
Fusão de métricas		
Média	0,86	0,84
Média ponderada	0,91	0,89
Máximo	0,71	0,72
Mínimo	0,65	0,64
Mediana	0,85	0,83
Média geométrica	0,79	0,81

Os resultados obtidos ainda mostram um aumento significativo, com um crescimento de cerca de 13% para o melhor resultado, indo de 0,80 para 0,91 na correlação linear (LCC) e de 0,79 para 0,89 na correlação monotônica (SROCC), em documentos técnicos ao utilizar a técnica de divisão e conquista de componentes nesse par de OCR e base de dados.

4.3 Estudos complementares

Essa seção detalha dois experimentos extras para confirmar se a generalização de qualidade das técnicas de DIQA utilizadas entre bases de dados, assim como verificar o impacto do *data augmentation* nos três modelos treinados.

4.3.1 Cross Dataset

Para a validação entre bases de dados foi feito um pequeno experimento com os modelos que aprenderam a prever a qualidade dos componentes de tabela e carimbo do motor de OCR Tesseract na base de dados StaVer, utilizando os mesmos para tentar prever a qualidade dos componentes de tabela e carimbo do motor de OCR Paddle na base de dados FUNSD. A tabela 15 demonstra os resultados de fazer a troca de um modelo que foi treinado em um contexto/base de dados com um OCR específico para outro totalmente diferente. É possível notar que, embora os padrões de qual o melhor e o pior modelos continuem o mesmo, existe uma diminuição drástica nos coeficientes de correlação, o que implica que os métodos de DIQA utilizados não tem uma generalização entre contextos satisfatória.

Tabela 15: Correlação entre predição-valor real - <i>Cross dataset</i>		
Técnica	Base	Tabela
Transfer-DIQA	0,56	0,60
Transfer-Efficient-DIQA	0,61	0,68
Attention-DIQA	0,54	0,52

4.3.2 Data Augmentation

Para checar o impacto de *data augmentation* nos modelos de aprendizado foi refeito o teste da tabela 5, olhando a correlação entre as técnicas de DIQA com e sem fusão usando o Tesseract na base de dados StaVer, porém foi removido o *data augmentation*, deixando somente 960 imagens de treino no mesmo. A tabela 16 demonstra que há uma leve de 2% para 3% na maioria das técnicas analisadas, mostrando que o *data augmentation* tem um impacto positivo no treinamento, porém não tão grande.

Tabela 16: Resultados de qualidade DIQA - Par Tesseract e Staver - Sem *data augmentation*

Técnica	LCC	SROCC
Transfer-DIQA sem fusão	0,77	0,80
Transfer-Efficient-DIQA sem fusão	0,82	0,85
Attention-DIQA sem fusão	0,68	0,64
Fusão de métricas		
Média	0,89	0,87
Média ponderada	0,92	0,89
Máximo	0,74	0,75
Mínimo	0,66	0,67
Mediana	0,88	0,86
Média geométrica	0,83	0,85

4.4 Exemplo qualitativo

As figuras 20, 21, 22 e 23 se referem a uma imagem de teste específica e seus componentes tabela, carimbo e base. Cada imagem tem uma métrica de saída de OCR real, neste caso provida do PaddleOCR, e também um *Recognition Score* provido pelo método Transfer-DIQA. Utilizando a rede Transfer-DIQA diretamente na imagem original sem utilizar o método de segmentação e fusão proposto neste trabalho gera um valor de qualidade predito de 0,72 e a diferença entre a qualidade predita e a qualidade real de 0,18. Utilizando a segmentação e a fusão com média ponderada propostas neste trabalho é possível obter um valor de qualidade predito de 0,88 e a diferença entre o valor real e o valor predito é de somente 0,02. Este exemplo mostra como o método proposto tem melhoras significativas para estes tipos de documento.

free research
culture conference

Matthias Werner
Bachstraße 184
09626 Kronpe
Tel: 0375/435-7810
Fax: 0375/535435-7810
E-Mail: E: 0913304026

Service IF - Bureau TIE/OGP

Matthias Werner, Bachstraße 184, 09626 Kronpe

Jan Weber
Jahnstraße 94
83280 Stadt Wehlen

Ihre Bestellung vom
22. Juni 1997

Datum
15. November 2010

LIQUIDATION

Sehr geehrter Herr Weber,

Bitte überweisen Sie den folgenden Rechnungsbeitrag innerhalb von 28 Tagen auf das unten angegebene Konto.

Anzahl	Beschreibung	Einzelpreis	Gesamtpreis
1	Versandkosten	9,19	9,19 €
26	50m Kabeltrommel, Überspannungsschutz	25,99	675,74 €
27	6x Steckdose	5,99	161,73 €
18	4x Steckdose	4,99	89,82 €
Gesamtsumme			936,48 €
inkl. 19% MwSt			149,52 €

Es grüßt Sie
Matthias Werner

Gesellschafter:
Matthias Werner
UID-Nr.
DE 6913304026

Bankverbindung:
Konto 118 781
Sparkasse Kitzingen
BLZ 950 902 92

Geschäftsführer:
Matthias Werner
Gerichtsstand:
Sankt Andreasberg

Figura 20: Documento original. *Recognition Score*: 0,90

Anzahl	Beschreibung	Einzelpreis	Gesamtpreis
1	Versandkosten	9,19	9,19 €
26	50m Kabeltrommel, Überspannungsschutz	25,99	675,74 €
27	6x Steckdose	5,99	161,73 €
18	4x Steckdose	4,99	89,82 €
Gesamtsumme			936,48 €
inkl. 19% MwSt			149,52 €

Figura 21: Segmento "tabela" do documento. *Recognition Score*: 0,85

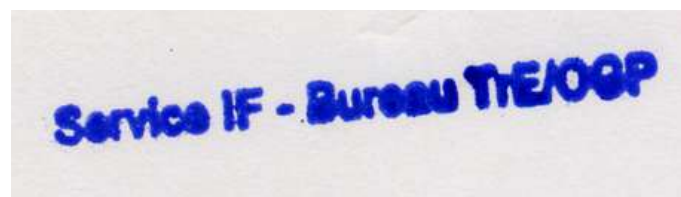


Figura 22: Segmento "carimbo" do documento. *Recognition Score*: 0,62

		 Matthias Werner Bachstraße 184 19626 Krempe Telefon: 0379 5759 425-7810 Fax: 0379 5759 425-7810 E-Mail: EW@F13M9076
Matthias Werner, Bachstraße 184, 19626 Krempe		
Jan Weber Jahnstraße 94 83280 Stadt Wehlen		
Ihre Bestellung vom 22. Juni 1997		Datum 15. November 2010
LIQUIDATION		
Sehr geehrter Herr Weber,		
Bitte überweisen Sie den folgenden Rechnungsbeitrag innerhalb von 28 Tagen auf das unten angegebene Konto.		
Es grüßt Sie Matthias Werner		
Gesellschafter: Matthias Werner UStId-Nr. DE 6913369676	Bankverbindung: Konto 118 781 Sparkasse Kasselburg BLZ 950 092 92	Geschäftsführer: Matthias Werner Gerichtsaussch. Saskia Andrusberg

Figura 23: Segmento "base" do documento. *Recognition Score: 0,95*

5 CONCLUSÃO

Neste trabalho foi colocado a estrutura para um modelo de análise de qualidade de imagens para documentos técnicos, onde foram feitos experimentos que mostram a eficácia da técnica de divisão e conquista durante o processo de DIQA, com o melhor resultado obtendo 96% de correlação linear entre a qualidade real e a qualidade predita, uma melhora de 14% comparado as técnicas do estado da arte.

Além disso foi levantado algumas peculiaridades sobre técnicas do estado da arte se comporta em diferentes contextos de documentos técnicos e motores de OCR, concluindo que algumas técnicas como CG-DIQA podem ser melhores em contextos sem formas complexas, ou que as técnicas de transferência de aprendizado tendem a serem melhores do que outras técnicas de DIQA para documentos técnicos.

Também foi analisado uma pequena melhora no estado da arte de DIQA utilizando *data augmentation*, porém essa melhora ainda não é algo totalmente concreto, visto que a quantidade de imagens utilizadas no treinamento não era tão grande e a melhora que é observada pela adição de *data augmentation* fica perto de 3%, um valor pequeno.

Outros experimentos foram feitos testando hipóteses para explicar a necessidade/viabilidade do método proposto, analisando a acurácia de motores OCR dependendo do componente, como métodos de DIQA atuais não se comportam bem em documentos técnicos complexos e métodos de detecção de tabelas e carimbos e sua robustez a degradação.

Algumas limitações existentes no trabalho contam com a utilização de algumas técnicas com redes defasadas como *backbone*, como a AlexNet. As decisões de utilizar as mesmas se dão pelo fato da área de DIQA não ter diversos trabalhos com tecnologias atualizadas, é como o trabalho era mais do que somente a utilização de DIQA foi necessário colocar um limite em quanto dessas técnicas específicas seriam utilizadas ou atualizadas para o resto do estado da arte de redes em visão computacional.

Outra limitação é o número de imagens pequeno, dado pelo fato de que *datasets* específicos que contêm todos os elementos necessários (texto, anotação de tabela e anotação de carimbo) não existem na literatura, fazendo com que seja necessário a anotação manual dos mesmos. Aumentar e diversificar bases de dados daria um panorama melhor das

vantagens e desvantagens do trabalho, assim como possibilitando a utilização de redes que necessitam de uma quantidade maior de informação para serem treinadas de forma satisfatória, como Transformers.

Outros métodos de OCR como o EasyOCR[39], funções de fusão e métricas de saída de OCR também poderiam ser adicionadas, melhorando a robustez do estudo e mostrando se existe algum problema nos resultados com a variação desses parâmetros.

Uma última adição poderia ser de componentes para a fusão. O *dataset* FUNSD tem anotações de formulários, o que são um componente de documentos, porém o mesmo não foi considerado no trabalho por questões de incompatibilidade desse componente com outras bases de dados, mas o mesmo poderia ser incluído para melhorar os resultados obtidos. O mesmo pode ser dito para outros componentes, como logos, áreas de texto manuscrito e gráficos.

REFERÊNCIAS

- [1] Al-Ani, A. and Deriche, M. (2002). A new technique for combining multiple classifiers using the dempster-shafer theory of evidence. *Journal of Artificial Intelligence Research*, 17:333–361.
- [2] Baek, Y., Nam, D., Park, S., Lee, J., Shin, S., Baek, J., Lee, C. Y., and Lee, H. (2020). Cleval: Character-level evaluation for text detection and recognition tasks. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*, pages 564–565.
- [3] Banerjee, S. and Lavie, A. (2005). Meteor: An automatic metric for mt evaluation with improved correlation with human judgments. In *Proceedings of the acl workshop on intrinsic and extrinsic evaluation measures for machine translation and/or summarization*, pages 65–72.
- [4] Blando, L. R., Kanai, J., and Nartker, T. A. (1995). Prediction of ocr accuracy using simple image features. In *Proceedings of 3rd International Conference on Document Analysis and Recognition*, volume 1, pages 319–322. IEEE.
- [5] Blecher, L., Cucurull, G., Scialom, T., and Stojnic, R. (2023). Nougat: Neural optical understanding for academic documents. *arXiv preprint arXiv:2308.13418*.
- [6] Chen, G.-H., Yang, C.-L., and Xie, S.-L. (2006). Gradient-based structural similarity for image quality assessment. In *2006 international conference on image processing*, pages 2929–2932. IEEE.
- [7] Chlap, P., Min, H., Vandenberg, N., Dowling, J., Holloway, L., and Haworth, A. (2021). A review of medical image data augmentation techniques for deep learning applications. *Journal of Medical Imaging and Radiation Oncology*, 65(5):545–563.
- [8] Daniel, W. (1990). *Applied Nonparametric Statistics*. Duxbury advanced series in statistics and decision sciences. PWS-KENT Pub.

- [9] Du, Y., Li, C., Guo, R., Cui, C., Liu, W., Zhou, J., Lu, B., Yang, Y., Liu, Q., Hu, X., et al. (2021). Pp-ocrv2: Bag of tricks for ultra lightweight ocr system. *arXiv preprint arXiv:2109.03144*.
- [10] Forczmański, P., Smoliński, A., Nowosielski, A., and Małecki, K. (2020). Segmentation of scanned documents using deep-learning approach. In *Progress in Computer Recognition Systems 11*, pages 141–152. Springer.
- [11] Galton, F. (1877). Typical Laws of Heredity. , 15(388):492–495.
- [12] Groleau, A., Chee, K. W., Larson, S., Maini, S., and Boarman, J. (2023). Augraphy: A data augmentation library for document images. In *International Conference on Document Analysis and Recognition*, pages 384–401. Springer.
- [13] Huang, Z., Chen, K., He, J., Bai, X., Karatzas, D., Lu, S., and Jawahar, C. (2019). Icdar2019 competition on scanned receipt ocr and information extraction. In *2019 International Conference on Document Analysis and Recognition (ICDAR)*, pages 1516–1520. IEEE.
- [14] Jaderberg, M., Simonyan, K., Vedaldi, A., and Zisserman, A. (2014). Synthetic data and artificial neural networks for natural scene text recognition. *arXiv preprint arXiv:1406.2227*.
- [15] Jaume, G., Ekenel, H. K., and Thiran, J.-P. (2019). Funsd: A dataset for form understanding in noisy scanned documents. In *2019 International Conference on Document Analysis and Recognition Workshops (ICDARW)*, volume 2, pages 1–6. IEEE.
- [16] Jin, L., Cho, S., Liu, T.-J., Egiazarian, K., and Kuo, C.-C. J. (2012a). Performance comparison of decision fusion strategies in bmmf-based image quality assessment. In *Proceedings of The 2012 Asia Pacific Signal and Information Processing Association Annual Summit and Conference*, pages 1–4. IEEE.
- [17] Jin, L., Egiazarian, K., and Kuo, C.-C. J. (2012b). Perceptual image quality assessment using block-based multi-metric fusion (bmmf). In *2012 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 1145–1148. IEEE.
- [18] Kanopoulos, N., Vasanthavada, N., and Baker, R. L. (1988). Design of an image edge detection filter using the sobel operator. *IEEE Journal of solid-state circuits*, 23(2):358–367.
- [19] Karpinski, R. and Belaid, A. (2018). Zonemapalt: An alternative to the zonemap metric for zone segmentation and classification. In *2018 13th IAPR International Workshop on Document Analysis Systems (DAS)*, pages 357–362. IEEE.

- [20] Krizhevsky, A., Sutskever, I., and Hinton, G. E. (2012). Imagenet classification with deep convolutional neural networks. *Advances in neural information processing systems*, 25.
- [21] Kumar, J., Chen, F., and Doermann, D. (2012). Sharpness estimation for document and scene images. In *Proceedings of the 21st International Conference on Pattern Recognition (ICPR2012)*, pages 3292–3295. IEEE.
- [22] Ledinauskas, E., Ruseckas, J., Marozas, J., Karlauskas, K., Terentjevas, J., Mačijauskas, A., and Juršėnas, A. (2021). Automatic travel pattern extraction from visa page stamps using cnn models. *arXiv preprint arXiv:2112.00348*.
- [23] Lei, Z., Zhao, S., Song, H., and Shen, J. (2018). Scene text recognition using residual convolutional recurrent neural network. *Machine Vision and Applications*, 29:861–871.
- [24] Levenshtein, V. I. (1966). Binary Codes Capable of Correcting Deletions, Insertions and Reversals. *Soviet Physics Doklady*, 10:707.
- [25] Li, H., Zhu, F., and Qiu, J. (2018). Cg-diqa: No-reference document image quality assessment based on character gradient. In *2018 24th International Conference on Pattern Recognition (ICPR)*, pages 3622–3626. IEEE.
- [26] Li, H., Zhu, F., and Qiu, J. (2019). Towards document image quality assessment: a text line based framework and a synthetic text line image dataset. In *2019 International Conference on Document Analysis and Recognition (ICDAR)*, pages 551–558. IEEE.
- [27] Li, P., Peng, L., Cai, J., Ding, X., and Ge, S. (2017). Attention based rnn model for document image quality assessment. In *2017 14th IAPR International Conference on Document Analysis and Recognition (ICDAR)*, volume 1, pages 819–825. IEEE.
- [28] Liao, M., Wan, Z., Yao, C., Chen, K., and Bai, X. (2020). Real-time scene text detection with differentiable binarization. In *Proceedings of the AAAI conference on artificial intelligence*, volume 34, pages 11474–11481.
- [29] Lu, T. and Doms, A. (2019a). A deep transfer learning approach to document image quality assessment. In *2019 International Conference on Document Analysis and Recognition (ICDAR)*, pages 1372–1377. IEEE.
- [30] Lu, T. and Doms, A. (2019b). Towards content independent no-reference image quality assessment using deep learning. In *2019 IEEE 4th International Conference on Image, Vision and Computing (ICIVC)*, pages 276–280. IEEE.

- [31] Madhusudana, P. C., Birkbeck, N., Wang, Y., Adsumilli, B., and Bovik, A. C. (2022). Image quality assessment using contrastive learning. *IEEE Transactions on Image Processing*, 31:4149–4161.
- [32] Micenkov, B. and van Beusekom, J. (2011). Stamp detection in color document images. In *2011 International Conference on Document Analysis and Recognition*, pages 1125–1129. IEEE.
- [33] Miranker, W. and Pan, V. (1980). Methods of aggregation. *Linear Algebra and its Applications*, 29:231–257. Special Volume Dedicated to Alson S. Householder.
- [34] Olejniczak, K. and Šulc, M. (2022). Text detection forgot about document ocr.
- [35] Paliwal, S. S., Vishwanath, D., Rahul, R., Sharma, M., and Vig, L. (2019). Tablenet: Deep learning model for end-to-end table detection and tabular data extraction from scanned document images. In *2019 International Conference on Document Analysis and Recognition (ICDAR)*, pages 128–133. IEEE.
- [36] Papineni, K., Roukos, S., Ward, T., and Zhu, W.-J. (2002). Bleu: a method for automatic evaluation of machine translation. In *Proceedings of the 40th annual meeting of the Association for Computational Linguistics*, pages 311–318.
- [37] Peng, X., Cao, H., Subramanian, K., Prasad, R., and Natarajan, P. (2011). Automated image quality assessment for camera-captured ocr. In *2011 18th IEEE International Conference on Image Processing*, pages 2621–2624. IEEE.
- [38] Prasad, D., Gadpal, A., Kapadni, K., Visave, M., and Sultanpure, K. (2020). Cascadetabnet: An approach for end to end table detection and structure recognition from image-based documents. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition workshops*, pages 572–573.
- [39] Shi, B., Bai, X., and Yao, C. (2016). An end-to-end trainable neural network for image-based sequence recognition and its application to scene text recognition. *IEEE transactions on pattern analysis and machine intelligence*, 39(11):2298–2304.
- [40] Shi, J., Gao, P., and Smolic, A. (2023). Blind image quality assessment via transformer predicted error map and perceptual quality token. *IEEE Transactions on Multimedia*.
- [41] 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.

- [42] Smock, B., Pesala, R., and Abraham, R. (2022). Pubtables-1m: Towards comprehensive table extraction from unstructured documents. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 4634–4642.
- [43] Souza, A., Cheriet, M., Naoi, S., and Suen, C. Y. (2003). Automatic filter selection using image quality assessment. In *Seventh International Conference on Document Analysis and Recognition, 2003. Proceedings.*, pages 508–512. IEEE.
- [44] Wang, Z. and Bovik, A. C. (2002). A universal image quality index. *IEEE signal processing letters*, 9(3):81–84.
- [45] Wang, Z., Bovik, A. C., Sheikh, H. R., and Simoncelli, E. P. (2004). Image quality assessment: from error visibility to structural similarity. *IEEE transactions on image processing*, 13(4):600–612.
- [46] Xu, J., Ye, P., Li, Q., Liu, Y., and Doermann, D. (2016). No-reference document image quality assessment based on high order image statistics. In *2016 IEEE International Conference on Image Processing (ICIP)*, pages 3289–3293. IEEE.
- [47] Zhong, X., Tang, J., and Yepes, A. J. (2019). Publaynet: largest dataset ever for document layout analysis. In *2019 International conference on document analysis and recognition (ICDAR)*, pages 1015–1022. IEEE.

A HIPÓTESES

A.1 Validação

O primeiro experimento feito teve o intuito de validar as hipóteses iniciais do trabalho, sendo elas:

1. OCR sofre perda de acurácia de forma diferente dependendo do componente(tabelas, carimbo e texto).
2. Os métodos de DIQA atuais não se comportam bem em documentos técnicos complexos.
3. Os métodos de detecção de tabelas e carimbos são robustos a degradação.

A.1.1 Hipótese 1

Para analisar a acurácia do OCR foi utilizado o *dataset* StaVer[32] em conjunto com as métricas CER e ZoneMapAlt[19], duas métricas que medem o erro do OCR. Foram utilizadas as degradações mencionadas anteriormente para fazer a comparação dos resultados do OCR antes e depois das mesmas. Para cada imagem os componentes a serem estudados(tabelas, carimbo e texto) foram separados e processados por um sistema de OCR, tendo como resultado final a média das duas métricas. A métrica de CER foi utilizada pois ela é a métrica padrão para medir a acurácia do reconhecimento de um sistema OCR e a métrica ZoneMapAlt foi utilizada porque ela foi criada para medir exclusivamente a acurácia da detecção de texto de um sistema de OCR, sendo ter uma métrica separada para cada é importante para verificar se as degradações podem afetar as duas etapas de um sistema OCR(detecção de texto e reconhecimento de texto) de forma diferente. O sistema de OCR utilizado foi o Tesseract[41] e sua configuração pode ser vista na seção 3.3. A tabela 17 contém os resultados levantados para a métrica de CER e a tabela 18 contém os resultados levantados para a métrica de ZoneMapAlt, demonstrando as métricas de saída antes e depois da degradação.

Como pode ser visto nas tabelas as métricas pioram com a degradação das imagens, mas as mesmas aparentam piorar mais os carimbos e as tabelas, em especial a

Tabela 17: CER do OCR (valor menor é melhor)

Componente	CER antes	CER depois
Tabela	0,34	2,51
Carimbo	8,98	17,35
Base	0,19	1,44

Tabela 18: ZoneMapAlt do OCR (valor menor é melhor)

Componente	ZoneMapAlt antes	ZoneMapAlt depois
Tabela	5,91	21,02
Carimbo	10,02	34,70
Base	1,26	5,30

segmentação das mesmas sofre muito, perdendo quase toda a informação original.

A.1.2 Hipótese 2

Para analisar as técnicas de DIQA atuais foram executados os dois métodos descritos na seção 3.5 na base de dados StaVer[32], separando cada um de seus componentes e testando cada um separadamente. Estes métodos tem como intuito prever a qualidade do OCR utilizando a métrica de CER. As tabelas 19, 20 e 21 mostram os resultados obtidos para cada um dos componentes.

Como pode ser visto em cada tabela a predição de CER chega razoavelmente perto do CER correto quando testada em imagens onde tabelas e carimbos são retirados, no componente "base", mas os resultados levando em conta seus componentes não são tão satisfatórios, especialmente no caso da RNN[27].

A.1.3 Hipótese 3

Para testar os métodos de detecção dos componentes após a degradação foi utilizado a base de dados StaVer[32] dividida em duas partes, com e sem degradação. Após isso foi feito a execução de redes neurais para a detecção dos componentes nas duas partes do *dataset*. Os resultados de detecção de tabelas podem ser visto na tabela 22 e os resultados de detecção de carimbos podem ser vistos na tabela 23.

Tabela 19: Predição dos métodos de DIQA atuais em tabelas

Método	CER Predito	CER certo
Lu and Doods [29]	3,18	2,51
Li et al. [27]	8,22	2,51

Tabela 20: Predição dos métodos de DIQA atuais em carimbos

Método	CER Predito	CER certo
Lu and Dooms [29]	14,92	17,35
Li et al. [27]	7,37	17,35

Tabela 21: Predição dos métodos de DIQA atuais em documento base

Método	CER Predito	CER certo
Lu and Dooms [29]	2,38	1,44
Li et al. [27]	3,75	1,44

Tabela 22: Detecção de tabelas

Método	Acurácia original	Acurácia com degradação
Smock et al. [42]	99,30%	93,27%
Paliwal et al. [35]	96,55%	88,71%
Prasad et al. [38]	91,51%	74,28%

Tabela 23: Detecção de carimbos

Método	Acurácia original	Acurácia com degradação
Forczmański et al. [10]	98,53%	89,20%
Ledinauskas et al. [22]	81,11%	69,37%

Como pode ser visto nas tabelas os resultados obtidos em modelos de redes neurais do estado da arte para a detecção do componente de tabela não são tão afetados quanto o OCR, enquanto a detecção de carimbos tem uma queda significativa na detecção, embora o mesmo ainda apresente quase 90% de acurácia. Este teste teve como intuito saber se o processo de DIQA proposto seria viável em uma aplicação real, visto que é necessário saber a localização dos componentes de tabela e carimbo antes de fazer o mesmo.