Artistic Text Recognition

Experiência Criativa: Projeto Transformador II
Bacharelado em Ciência da Computação – PUCPR

Turma U – Equipe 7

Guilherme Henrique Eduardo de Lara Peres, Henrique Anderle Schulz, Lucas Azevedo Dias,
Pedro Lucas Ghezzi Bittencourt, Rafaela de Miranda

{guilherme.peres, henrique.schulz, dias.azevedo, pedro.bittencourt, r.miranda2}@pucpr.edu.br

I. Descrição do Projeto

O reconhecimento de texto artístico (*artistic text recognition*) tem ganhado relevância em aplicações como publicidade, design gráfico e mídias digitais. Contudo, essa tarefa permanece pouco explorada na literatura científica e apresenta desafios significativos para sistemas automáticos de reconhecimento, tais como fontes estilizadas, efeitos visuais, sobreposição de caracteres e fundos complexos [1]. Esses fatores tornam o problema mais desafiador do que o reconhecimento de texto em contextos convencionais.

A competição ICDAR 2024 – Artistic Text Recognition destacou essas dificuldades ao propor a tarefa de transcrever corretamente textos artísticos a partir do conjunto de dados WordArt-V1.5, composto por 12.000 imagens divididas igualmente entre treino e teste [1]. Esse *dataset* contempla uma grande diversidade de estilos, incluindo cartazes, cartões e composições tipográficas elaboradas, representando um cenário desafiador para modelos atuais de OCR (*Optical Character Recognition*).

O objetivo central deste projeto é avaliar diversos modelos capazes de reconhecer textos artísticos e entender suas nuances mesmo em condições adversas impostas pelo estilo visual das imagens. Especificamente, busca-se: (i) averiguar a acurácia de reconhecimento de palavras pelos modelos do estado da arte através da *Word Recognition Accuracy* (WRA), métrica oficial da competição; (ii) avaliar a eficácia de técnicas do estado da arte que lidem com fontes ornamentadas, sobreposição de caracteres e ruídos visuais; e (iii) comparar os resultados obtidos do estado da arte e consolidar os avanços na área.

II. Materiais e Métodos

Para a realização deste projeto, utilizaremos como base o conjunto de dados WordArt-V1.5, disponibilizado no contexto da competição ICDAR 2024 [1]. Esse *dataset* contém 12.000 imagens de textos artísticos divididas igualmente em treino e teste, abrangendo diferentes estilos tipográficos, composições gráficas e variações visuais. A riqueza e a diversidade dessas imagens representam um desafio relevante para os métodos tradicionais de reconhecimento de texto, que geralmente são projetados para cenários mais regulares.

A Figura 1 apresenta amostras representativas do conjunto WordArt-V1.5, evidenciando a variabilidade estilística e os principais desafios visuais.

Figura 1. Exemplos de amostras do conjunto de dados WordArt-V1.5, com diferentes estilos tipográficos e complexidades visuais.

No que se refere às arquiteturas, serão investigados três modelos recentes do estado da arte em reconhecimento de texto.

Um deles é o SVTRv2, um modelo baseado na arquitetura CTC (Connectionist Temporal Classification), que utiliza uma estratégia de redimensionamento multi-tamanho (MSR) e um módulo de orientação semântica (SGM) para lidar com as limitações do CTC, como irregularidades no texto e contexto linguístico, se saindo melhor que modelos STR convencionais e até mesmo modelos EDTR (Encoder-decoder-based Text Recognition) [2].

O segundo modelo é o ViTSTR, um modelo com foco na velocidade e eficiência, que utiliza uma arquitetura simplificada utilizando *vision transformers* (ViT), com camadas e parâmetros reduzidos, resultando em uma melhora, comparado

Tabela I
RESUMO DE MATERIAIS E MÉTODOS.

Materiais	Descrição	Observações
Conjunto de Dados	WordArt-V1.5 (ICDAR 2024)	12.000 imagens de texto artístico com grande diversidade de estilos e ruídos visuais.
Modelos	FAST (Fourier Aided Scene Text Recognition), PARSeq (Permuted Autoregressive Sequence Model)	Métodos recentes do estado da arte em OCR, adaptados para lidar com estilização intensa.
Técnicas	Transfer Learning, Data Augmentation	Aumento de dados com distorções geométricas, ruído visual e variação tipográfica.
Linguagens / Bibliotecas	Python, PyTorch, OpenCV, scikit-learn	Frameworks modernos para visão computacional e aprendizado profundo.
Hardware / Recursos	Google Colab (GPU Tesla T4), Desktop com GPU RTX 3060	Recursos disponíveis para prototipagem e treinamento de modelos.

ao modelo TRBA (TPS-ResNet-BiLSTM-Attention), de até 2.5x na velocidade de inferência, com uma perda de aproximadamente 3% na acurácia na versão "*tiny*". No entanto o modelo permite outras configurações que podem ser usadas para equilibrar o desempenho [3].

O terceiro modelo de interesse é o **PARSeq** (Permuted Autoregressive Sequence Model), que formula o reconhecimento de texto como um processo autoregressivo com permutações, obtendo resultados competitivos em benchmarks complexos [4]. Os três métodos fornecem pontos de partida promissores para lidar com os desafios presentes no WordArt-V1.5.

Além disso, o projeto fará uso de técnicas de *data augmentation* específicas para textos artísticos, incluindo simulação de ruídos gráficos, distorções geométricas e variações de cor e estilo. Estratégias de *transfer learning* também serão exploradas, aproveitando modelos pré-treinados em grandes bases de dados e adaptando-os ao domínio específico de texto artístico.

As implementações serão realizadas majoritariamente em **Python**, utilizando bibliotecas como PyTorch, OpenCV e scikit-learn. O ambiente de desenvolvimento incluirá Google Colab para prototipagem e uma máquina local equipada com uma unidade de processamento gráfico (*GPU*) RTX 3060 Mobile, garantindo capacidade computacional adequada para o processamento dos modelos.

III. ETAPAS DO PROJETO E MARCOS FÍSICOS

O desenvolvimento do projeto será organizado em etapas sequenciais, cada uma acompanhada de marcos que permitirão avaliar seu progresso e assegurar a qualidade das entregas. A primeira etapa consiste na elaboração da revisão de literatura a partir do levantamento do estado da arte prévio, cujo objetivo é consolidar o referencial teórico necessário para sustentar as escolhas metodológicas e experimentais do estudo. A conclusão desta fase será evidenciada pela entrega parcial do texto correspondente, que será submetido à validação.

Na sequência, será realizada a execução experimental dos modelos propostos, abrangendo três implementações distintas. Cada modelo será implementado e executado individualmente, sendo a validação dessa etapa garantida pela disponibilização dos resultados obtidos e do respectivo código-fonte, acessível por meio do Google Colab.

Após a execução dos modelos, os resultados gerados serão incorporados ao manuscrito, assegurando que a análise crítica e a discussão estejam fundamentadas em evidências experimentais. Essa fase será validada mediante a entrega parcial do documento atualizado, permitindo o acompanhamento da coerência entre os dados obtidos e a argumentação apresentada.

Por fim, será realizada a entrega final do projeto, contemplando a versão definitiva do manuscrito e a disponibilização completa dos códigos-fonte utilizados no estudo. Essa etapa representará a consolidação de todo o trabalho desenvolvido, garantindo a transparência e a rastreabilidade dos procedimentos adotados, em conformidade com os objetivos definidos no planejamento inicial.

IV. CRONOGRAMA

Tabela II
CRONOGRAMA DO PROJETO

Atividade	Sem. 1–2	Sem. 3–4	Sem. 5–6	Sem. 7–8	Sem. 9–11 (até 09/11)
Levantamento do estado da arte	X				
Revisão de literatura	X	X			
Execução do modelo 1		X			
Execução do modelo 2		X	X		
Execução do modelo 3			X		
Incorporação dos resultados				X	
Entrega final					X

As semanas são contadas a partir da entrega deste planejamento até a entrega final em 09/11.

REFERÊNCIAS

[1] "Icdar 2024 competition on artistic text recognition (wordart-v1.5)," https://sites.google.com/view/icdar-2024-competition-wordart, acessado em: 26 ago. 2025.

[2] H. X. C. J. Y.-G. J. Yongkun Du, Zhineng Chen, "Svtrv2: Ctc beats encoder-decoder models in scene text recognition," in *International Conference on Computer Vision (ICCV)*, 2025.

[3] R. Atienza, "Vision transformer for fast and efficient scene text recognition," in *ICDAR2021*, 2021.

[4] D. Bautista and R. Atienza, "Parseq: Permuted autoregressive sequence models for text recognition," in *International Conference on Pattern Recognition (ICPR)*, 2022.