Um keylogger acústico para ataque remoto usando aprendizado profundo

Palavras-Chave: Aprendizado profundo, Keylogging, Processamento de Sinal

Autores(as):

Maria Eduarda Xavier Messias – FEM

Prof. Dr. Josué Labaki (orientador) – FEM

1. INTRODUÇÃO:

Este trabalho tem como objetivo apresentar uma abordagem inovadora para o reconhecimento de teclas em um teclado semi mecânico, empregando o método de análise de espectrogramas e a técnica de Redes Neurais Convolucionais com máscara (Mask R-CNN). O propósito é desenvolver um sistema capaz de identificar as teclas pressionadas a partir dos sons capturados pelo microfone de um dispositivo móvel. A técnica de espectrograma mel-scaled é utilizada para transformar os sinais de áudio em representações visuais, as quais são processadas pela API detectron 2 do Facebook, permitindo a detecção e classificação das teclas em tempo real.

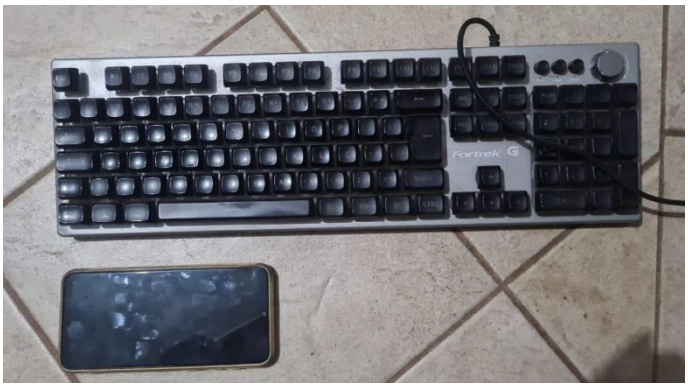
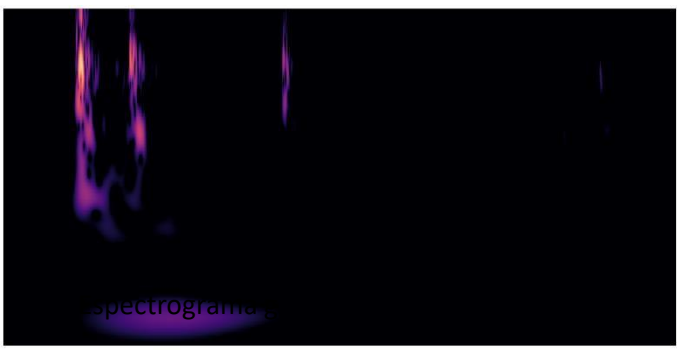


Fig.1: Fotografia do teclado mecânico e o celular Galaxy A30 utilizados para a gravação dos sons

2. METODOLOGIA:

O procedimento inicial consistiu na coleta dos sons gerados pelas teclas do teclado semi mecânico (Fig.1), utilizando-se um celular Samsung Galaxy A30 para capturar os áudios no formato WAV. Em seguida, aplicou-se a Transformada Wavelet



Contínua com a onda Morlet para gerar os espectrogramas. Essa técnica possibilita a análise das frequências e padrões temporais do som, transformando-os em informações visuais, como pode ser visto na figura (Fig.2)

Para o processamento dos espectrogramas, foi empregada a linguagem de programação Python, juntamente com bibliotecas específicas como Numpy, TensorFlow, Librosa, Data Frame e Pickle. O framework Detectron 2, desenvolvido pelo Facebook, foi utilizado para treinar e avaliar o modelo Mask R-CNN. Essa API simplifica a implementação de modelos de detecção de objetos e permite a exportação dos dados em formato COCO (Common Objects in Context), tornando a integração com o modelo de detecção mais eficiente.

O modelo selecionado para a detecção e segmentação das teclas foi o "mask_rcnn_R_101_DC5_3x". Essa arquitetura é uma extensão da Faster R-CNN, combinada com uma máscara de segmentação. Embora a Faster R-CNN seja uma rede neural convolucional projetada para detectar objetos em imagens, ela não fornece informações detalhadas sobre a forma precisa dos objetos detectados. Por isso, a máscara adicionada à Faster R-CNN é responsável por gerar uma máscara binária que delimita com precisão a região do objeto, permitindo uma segmentação mais detalhada e precisa das teclas.

3. RESULTADOS E DISCUSSÃO:

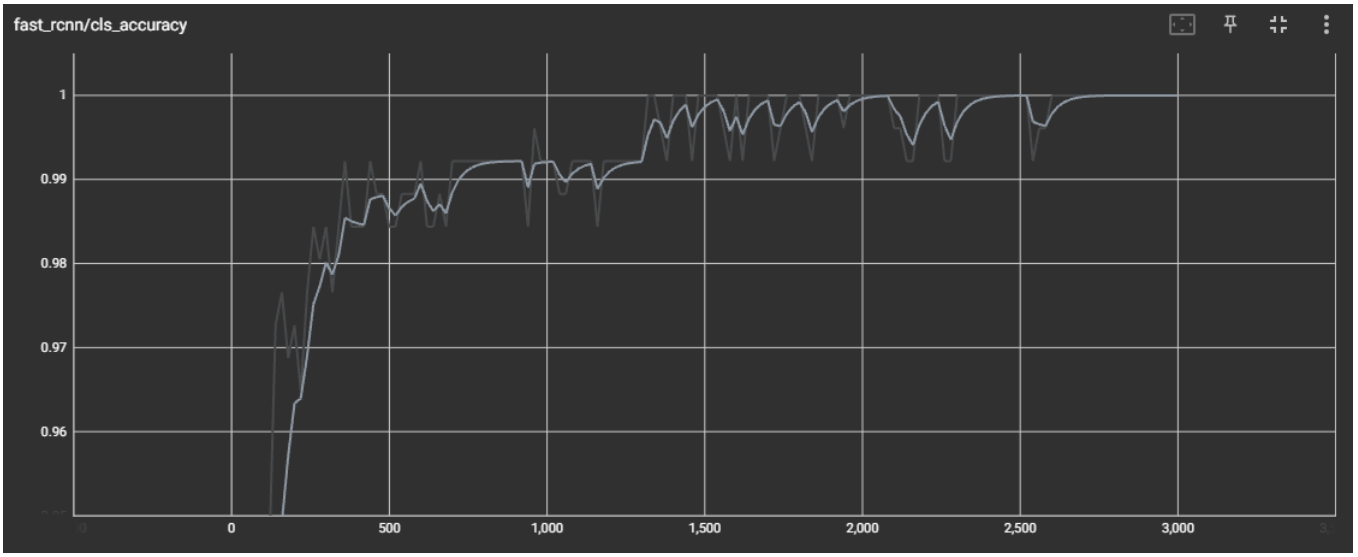


Fig.3: Gráfico da Progressão de Acurácia da rede, o gráfico de Iterações por acurácia

Após o treinamento e avaliação do modelo Mask R-CNN, utilizamos um conjunto de dados composto por 624 espectrogramas de teclas pressionadas, classificados em seis categorias: A, S, D, Q, W, E. A precisão geral do modelo na detecção e classificação das caixas delimitadoras das instâncias (bounding boxes) foi de 99% (Fig.3). Essa alta precisão indica

que o modelo foi capaz de identificar com extrema acurácia as regiões do espectrograma correspondentes às teclas pressionadas.

A métrica de taxa de falsos negativos apresentou um valor extremamente baixo, de apenas 0.001 (Fig.4). Isso significa que o modelo raramente deixou de detectar uma tecla pressionada, demonstrando sua capacidade excepcional de reconhecimento das teclas. A precisão na classificação das regiões positivas (foreground class accuracy) foi de 99.8%, reforçando ainda mais a robustez do modelo em identificar corretamente as classes de teclas.

Adicionalmente, analisamos as perdas durante o treinamento do modelo. A perda na regressão das coordenadas das caixas delimitadoras (loss box regression) obteve um valor de 0.05, indicando que o modelo aprendeu efetivamente a posicionar as caixas delimitadoras com precisão. A perda na classificação das regiões de interesse (loss RPN classification) foi de apenas 0.0007, evidenciando a eficiência do modelo em classificar as regiões de interesse como positivas ou negativas.

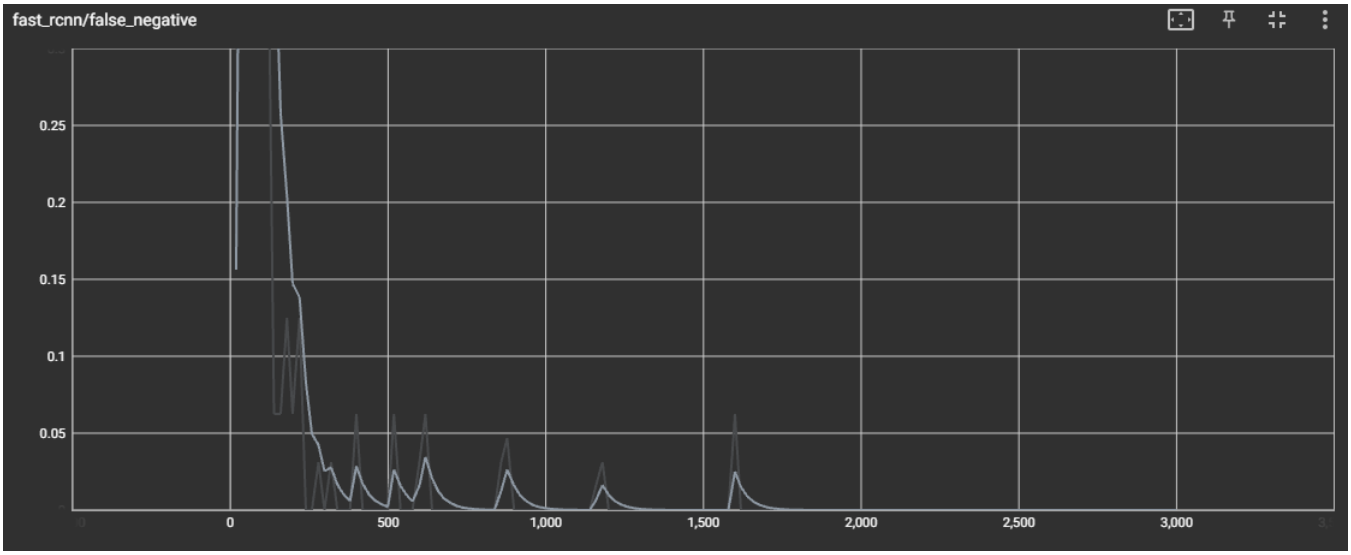


Fig.4: Gráfico da Progressão de Acurácia da rede para falsos negativos, o gráfico de Iterações por acurácia

4. CONCLUSÕES:

Este estudo proporcionou uma abordagem eficiente para o reconhecimento de teclas em um teclado semi mecânico, por meio da análise de espectrogramas e da técnica de Redes Neurais Convolucionais com máscara (Mask R-CNN). Os resultados obtidos revelaram a notável precisão do modelo na detecção e classificação das teclas pressionadas, com uma taxa de falsos negativos excepcionalmente baixa.

A utilização da API Detectron 2 do Facebook facilitou significativamente o treinamento e a avaliação do modelo, permitindo também a exportação dos dados em formato COCO. O

modelo "mask_rcnn_R_101_DC5_3x" destacou-se pela sua eficiência na segmentação precisa das teclas, o que representa um avanço na tecnologia de reconhecimento de teclas em dispositivos de entrada.

Espera-se que os resultados desta pesquisa possam ser aplicados em diversas áreas, como processamento de imagem e som, contribuindo para o desenvolvimento de sistemas mais eficientes e precisos. O estudo foi conduzido na FEM, sob orientação do Prof. Dr. Josué Labaki, e acredita-se que esses resultados contribuirão para a disseminação do conhecimento e inspirarão novas pesquisas na área de processamento de sinais de áudio e reconhecimento de padrões.

5. BIBLIOGRAFIA

Aktaş, Y. Ç. (8 de March de 2023). *Object Detection with Convolutional Neural Networks | by Yağmur Çiğdem Aktaş*. Fonte: Towards Data Science: https://towardsdatascience.com/object-detection-with-convolutional-neural-networks-c9d729eedc18

Grassin, C. (27 de 2019 de December). *Breaking Passwords with a Microphone*. Acesso em 8 de March de 2023, disponível em Charles' Labs: https://charleslabs.fr/en/project-Breaking+Passwords+with+a+Microphone

Lazebnik, L. (16 de Setembro de 2022). *Convolutional Neural Network Architectures: from LeNet to ResNet*. Fonte: Convolutional Neural Network Architectures: from LeNet to ResNet: https://slazebni.cs.illinois.edu/spring17/lec01_cnn_architectures.pdf

Magana, R. (s.d.). *The best machine learning model for binary classification*. Fonte: Ruslan Magana Vsevolodovna: https://ruslanmv.com/blog/The-best-binary-Machine-Learning-Model

Shenoy, K. (17 de January de 2021). *Keystroke Dynamics Analysis and Prediction — Part 1/2 (EDA) | by Kartik Shenoy*. Acesso em 6 de June de 2023, disponível em Towards Data Science: https://towardsdatascience.com/keystroke-dynamics-analysis-and-prediction-part-1-eda-3fe2d25bac04