Detecting Deepfake Audio with Spectrogram Analysis Using Convolutional Neural Networks: A Comprehensive Approach

A Dekhane Aishwarya
4 min read · Aug 1, 2024

Introduction

The proliferation of deepfake technology has created a pressing need for robust detection methods, particularly in the realm of audio and visual media. To address this challenge, we have developed a Convolutional Neural Network (CNN) model that analyzes spectrogram images to distinguish between authentic and deepfake audio recordings. Leveraging the power of deep learning, our model has achieved an impressive accuracy rate, demonstrating its effectiveness in identifying manipulated content.

Model Architecture and Data Preparation

The core of our approach lies in the CNN's ability to automatically learn and extract relevant features from spectrogram images. The model architecture consists of several layers designed to capture different levels of abstraction from the input data:

1. Convolutional Layers: Three convolutional layers are used, each with increasing filter sizes (32, 64, and 128). These layers apply convolution operations to extract features such as edges, textures, and more complex patterns. Each convolutional layer is followed by a ReLU activation function to introduce non-linearity and a max-pooling layer to reduce the spatial dimensions, thereby controlling overfitting and improving computational efficiency.

2. Fully Connected Layers: The flattened output from the final convolutional layer is passed through two fully connected layers. The first has 128 neurons and uses a ReLU activation function, while the second is the output layer with two neurons (one for each class: original and deepfake) and a softmax activation function, which provides probability scores for classification.

3. Regularization Techniques: To prevent overfitting, dropout is applied with a rate of 0.5, randomly setting half of the inputs to zero during training, which helps in regularizing the model.

Data Acquisition and Preprocessing

We curated a dataset comprising spectrogram images generated from both original and deepfake audio files. The deepfake audio samples were sourced from a collection of synthetically generated voice recordings, while the original samples were from genuine recordings. The images were

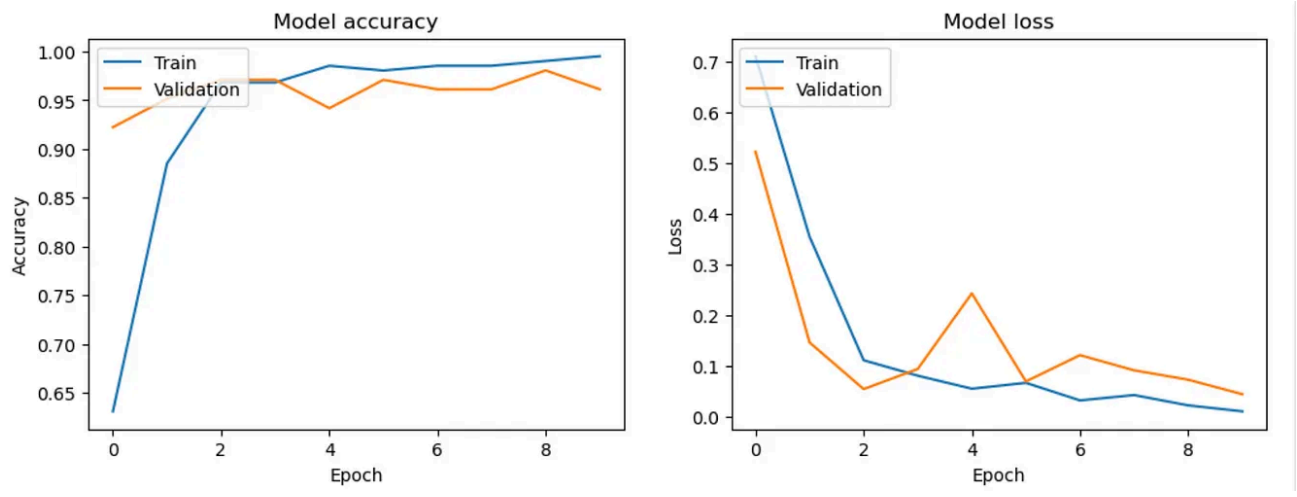
standardized to a consistent size of 128x128 pixels and normalized to a range of [0, 1] by dividing pixel values by 255.

The dataset was split into training and testing sets, with an 80–20 ratio. The training set was further divided to include a validation set, ensuring that the model's performance could be monitored on unseen data during the training process.

Model Training and Evaluation

The CNN model was trained for 10 epochs with a batch size of 32. The Adam optimizer was used for its efficiency in handling large datasets and adaptive learning rate capabilities. The model's loss was minimized using categorical crossentropy, suitable for our binary classification problem.

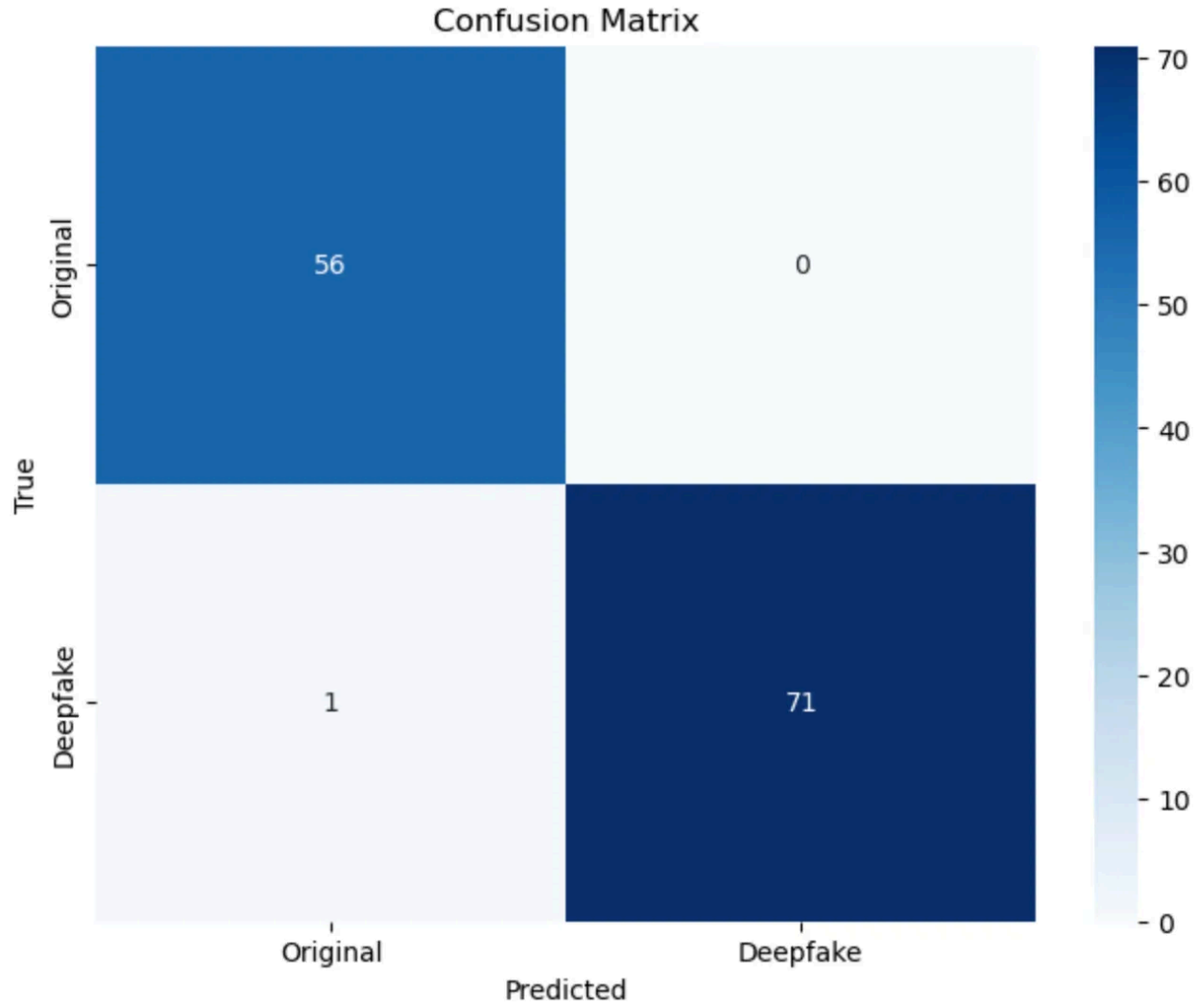
During training, both training and validation accuracy were closely monitored. The model demonstrated steady improvement, ultimately achieving a test accuracy of 98%. This high accuracy indicates the model's capability to distinguish between original and deepfake audio spectrograms with high precision.



Results and Analysis

The performance of the model was evaluated using a confusion matrix and a classification report, which provided detailed insights into precision, recall, and F1-scores for both classes. The confusion matrix revealed a low false positive and false negative rate, underscoring the model's reliability.

The visualization of the training process showed a clear convergence of both loss and accuracy metrics, indicating effective learning. The model's ability to generalize well to new data was further confirmed by the minimal gap between training and validation accuracies.



Conclusion

This project showcases the potential of CNNs in deepfake detection, specifically through the analysis of spectrogram images. The high accuracy achieved demonstrates the model's robustness and effectiveness in identifying manipulated audio content. As deepfake technology evolves, continuous improvements and adaptations of detection models will be crucial. Future work may involve exploring more complex architectures, incorporating additional data modalities, and addressing adversarial attacks to enhance detection resilience.

The successful deployment of this CNN model represents a significant step forward in the fight against deepfake audio, contributing to the broader effort to maintain the integrity of digital media.

	precision	recall	f1-score	support
Original	0.98	1.00	0.99	56
Deepfake	1.00	0.99	0.99	72
accuracy			0.99	128
macro avg	0.99	0.99	0.99	128
weighted avg	0.99	0.99	0.99	128

Deep Learning | Deepfakes | Artificial Intelligence | Machine Learning

A

Written by Dekhane Aishwarya

1 Follower · 2 Following

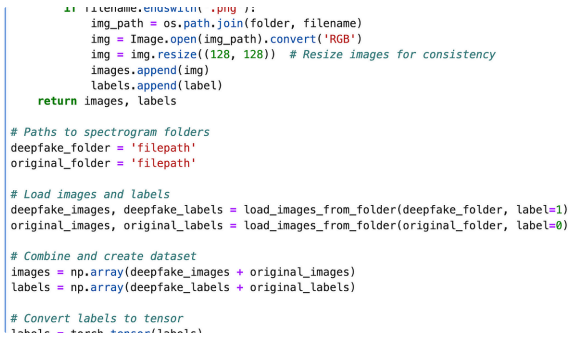
Edit profile

No responses yet

What are your thoughts?

Respond

More from Dekhane Aishwarya

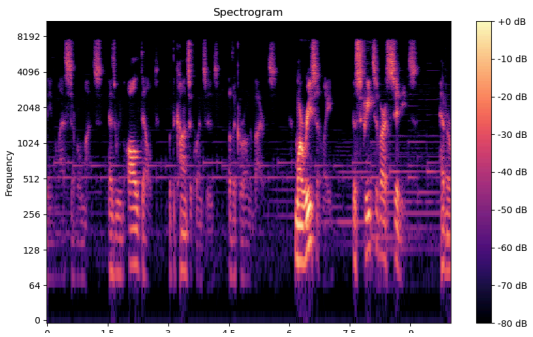


```
        if filename.endswith('.png'):
            img_path = os.path.join(folder, filename)
            img = Image.open(img_path).convert('RGB')
            img = img.resize((128, 128))  # Resize images for consistency
            images.append(img)
            labels.append(label)
    return images, labels

# Paths to spectrogram folders
deepfake_folder = 'filepath'
original_folder = 'filepath'

# Load images and labels
deepfake_images, deepfake_labels = load_images_from_folder(deepfake_folder, label=1)
original_images, original_labels = load_images_from_folder(original_folder, label=0)

# Combine and create dataset
images = np.array(deepfake_images + original_images)
labels = np.array(deepfake_labels + original_labels)

# Convert labels to tensor
```

A Dekhane Aishwarya

Understanding Activation Layers in Deepfake Detection Using a CNN

In the realm of machine learning, particularly in convolutional neural networks (CNNs),...



A Dekhane Aishwarya

Deepfake Detection Using Spectrogram Analysis: An...

In recent years, the rise of deepfake technology has presented significant...

Sep 29, 2024 8 Jul 15, 2024

Dekhane Aishwarya

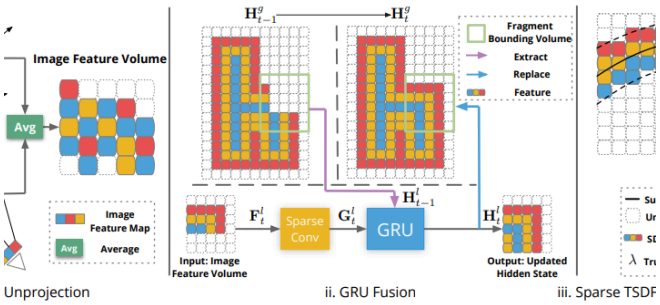
Utilizing Image Morphology for Enhanced Deepfake Detection in...

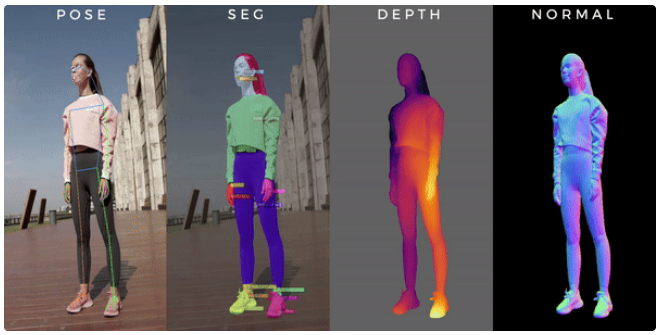
Deepfake detection has become a significant area of research due to the increasing...

Jul 23, 2024

See all from Dekhane Aishwarya

Recommended from Medium





The useless channel

Neural Recon: 4D reconstruction

NeuralRecon: The latest method for real-time 3D scene reconstruction from a monocular…

Aug 20, 2024

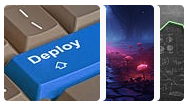
AI Papers Academy

Sapiens by Meta AI: Foundation for Human Vision Models

In this post we dive into Sapiens, a new family of computer vision models by Meta AI that…

Aug 26, 2024 130

Lists

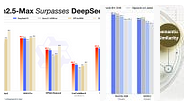


Predictive Modeling w/ Python

20 stories · 1809 saves

AI Regulation

6 stories · 679 saves

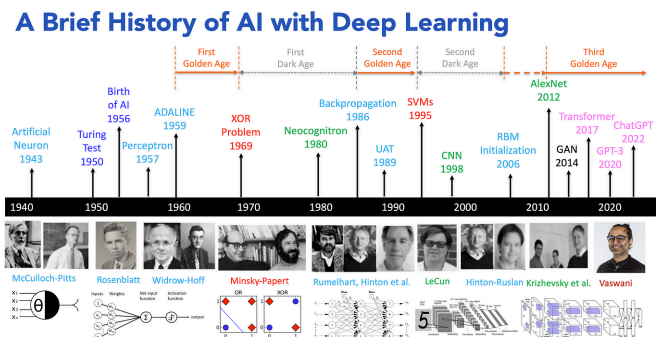


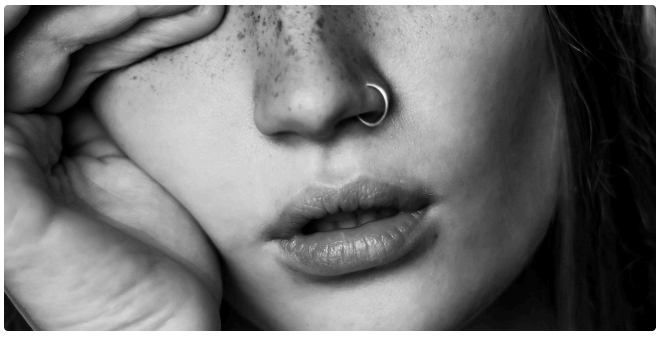
Natural Language Processing

1908 stories · 1566 saves

Practical Guides to Machine Learning

10 stories · 2182 saves





LM Po

A Brief History of AI with Deep Learning

Artificial intelligence (AI) and deep learning have seen remarkable progress over the pas...

Sep 1, 2024 433 8

Marko Briesemann

Emotion recognition with AI

Several libraries and models are available for emotion recognition from audio, analyzing...

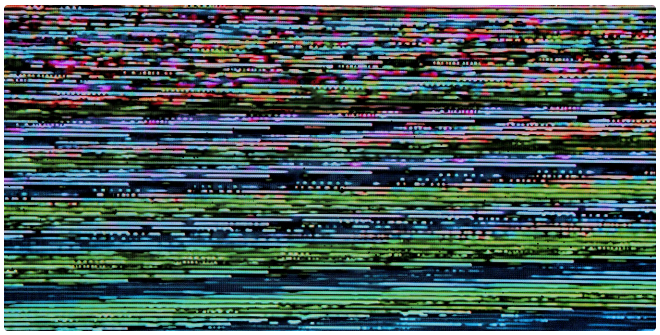
Nov 15, 2024 2

In Generative AI by Jim Clyde Monge

How To Install And Use DeepSeek R-1 In Your Local PC

Here's a step-by-step guide on how you can run DeepSeek R-1 on your local machine eve...

Jan 23 2.1K 47



In Towards Data Science by Piero Paialunga

Hands-on Generative Adversarial Networks (GAN) for Signal...

Here's how to build a generative Deep Learning model for Signal Processing in a fe...

Dec 27, 2022 558 6

See more recommendations