Devin Liu, Khiem Do

Gady Agam

CS512 Computer Vision

Oct 17, 2024

Deepfake Image Detection Using an Improved Dense CNN Architecture.

Option 2 - Modify a significant component of the paper

Problem Description:

With the recent rise and progression of AI technology, society finds itself face to face with a growing problem: deepfake content. Deepfakes allow for the creation of highly realistic but fake images, videos, or even audio, which can be used to spread misinformation, posing as a threat to personal reputations or even public opinion. Our project addresses this issue by focusing on improving the Dense CNN (D-CNN) architecture proposed by Yogesh Patel et al. for detecting deepfake images generated using Generative Adversarial Networks (GANs). We aim to enhance the speed and training efficiency of the model while maintaining accuracy.

Approach:

The goal is to create an improved mobileNet architecture to detect deepfake images. The model will extract deepfake image features using multiple convolutional layers as well as depthwise separable convolution layers and apply a binary classification to distinguish real images vs GANs deepfake generated images.

We will implement a series of depthwise separable 2D convolutions to extract features from the input image, with each subsequent block applying depthwise convolutions followed by pointwise convolutions to capture deeper features more efficiently. This approach reduces the number of parameters and computational complexity significantly compared to traditional convolutions, which makes the model faster. Then, we will apply batch normalization and average pooling layers to reduce dimensions and computational complexity while maintaining key features, helping to avoid overfitting. Throughout the network, we will use leaky ReLU as the activation function, allowing for better gradient flow and preventing vanishing gradients. Afterwards, the features are flattened into a one-dimensional array that is the input to a fully connected layer. Finally, we will implement a sigmoid function to predict the outcome of the image, giving us classification results. The model makes use of the Adam optimizer with binary cross-entropy loss, aiming to maximize classification accuracy. The model's performance will be evaluated using accuracy, precision, recall, and F1-scores.

Data (Source):

We will use a combination of real and deepfake image datasets for training and testing. For real images, we will use the CelebA and FFHQ Datasets as used by the paper, while deepfake images will be sourced from various GAN architectures, such as StyleGAN, StarGAN, AttGAN, and GDWCT. The training set will consist of 5000 real images, and 5000 deepfake images, 1000 from each deepfake dataset. 30% of the data will be reserved for testing. A subset of the training data will be used for validation to mitigate overfitting.

Paper:

Our project is based on the findings from the paper titled "An Improved Dense CNN Architecture for Deepfake Image Detection" by Yogesh Patel et al., published in 2023. In our project, we will aim to recreate the model they devised from scratch.

<u>Team Member Responsibilities:</u>

- Devin Liu: Implementation of the D-CNN model, focusing on the optimization and experimentation of hyperparameters. Responsible for gathering real and deepfake datasets.
- Khiem Do: Data preprocessing and validation. Responsible for implementing the data pipeline, conducting performance evaluations, and creating the final report.

Source

Patel, Y., Tanwar, S., Bhattacharya, P., Gupta, R., Alsuwian, T., Davidson, I. E., & Mazibuko, T. F. (2023). An improved dense CNN architecture for Deepfake Image detection. *IEEE Access*, *11*, 22081–22095. https://doi.org/10.1109/access.2023.3251417