**Assignment: Applying Autoencoders**

Weights & Biases Link: https://wandb.ai/usf-guardians/GoTG\_Assignment07\_AE?nw=nwuserprincepraveen

In recent years, the challenge of identifying individuals in low-light conditions has gained prominence due to its applications in security and surveillance. This project focuses on enhancing the identification process within a dataset of images of potential culprits, specifically targeting an ice cream thief within a shared household environment consisting of six roommates. The primary objective is to utilize a Convolutional Autoencoder (CAE) to reconstruct clearer images of faces from blurry, grainy footage captured by a fridge camera, enabling the identification of the thief. This task involves selecting a suitable dataset of facial images, preprocessing the data to improve the robustness of the model, and developing a CAE to perform the image denoising and reconstruction.

Reflecting on the business context of our project, it can be framed with the following theme. For instance, this project tackles a very frustrating issue that often arises in shared living environments. By revealing the identity of the ice cream thief, the project promotes accountability and can foster a sense of community among roommates. Businesses focusing on conflict resolution or community building could explore similar methods to create tools or programs facilitating better interpersonal relations among group members.

This report outlines the rationale for dataset selection, the preprocessing techniques implemented, the architecture design of the CAE, evaluation metrics applied, and a comprehensive discussion on the model's performance and the challenges encountered during its deployment.

**Dataset and Preprocessing**

**Dataset Overview**

[Dataset Link](https://www.kaggle.com/datasets/jessicali9530/lfw-dataset)

The **Labeled Faces in the Wild (LFW)** dataset is a well-known benchmark for evaluating algorithms in unconstrained face recognition. It was developed by researchers at the University of Massachusetts, Amherst, and features real-world, varied conditions such as lighting, expressions, and occlusions. This version of the dataset includes 13,233 images of 5,749 individuals, with 1,680 individuals having two or more distinct photos. The images were detected and centered using the Viola-Jones face detection algorithm, then processed using a deep-funneling alignment technique, which has shown improved performance in face verification tasks. The dataset contains both image files and supporting metadata.

**Image Data:**

* Format: .jpg files located in the structure lfw/person\_name/person\_name\_xxxx.jpg
* Dimensions: 250x250 pixels, scaled after detection and alignment
* Alignment: Images were deep-funneled for better facial alignment and recognition performance

**Metadata Files (10 total):**

* lfwallnames.csv: Lists all individuals and the number of images associated with each
* lfwreadme.csv: A comprehensive guide detailing how to use the dataset, metadata descriptions, and suggested train-test split strategies
* Additional CSVs: Support tasks like creating subsets, verifying image pairs, and labeling

For this assignment, we selected 6 “roommates” from the dataset with 30+ face images each. Furthermore, all images were standardized to 64x64 pixels to represent our potential ice cream thieves. This selection provided a focused dataset to train the CAE on different facial features.

**Dataset Preprocessing:**

1. **Image Standardization:** All selected images were standardized to a uniform size of 64x64 pixels. This consistent resolution ensured that the input dimensions for the CAE were uniform, which is essential for effective learning and processing.
2. **Augmentation Strategies:** Different preprocessing strategies were tested for their effectiveness on the CAE model's robustness. The augmentations were applied to introduce variation in the training data and help prevent overfitting. Each team member employed a distinct set of augmentations:

* **Prince**: Used no augmentations, maintaining a *noise factor of 0.3* with inputs resized and transformed to tensors for a cleaner but challenging learning environment.
* **Debjani**: Utilized horizontal flips, rotations, and color jitter with a *noise factor of 0.15*, providing balanced and realistic augmentations to improve generalization
* **Kathryn**: Applied more extensive augmentations, including cropping, flipping, rotation, affine transformations, and grayscale, with a *noise factor of 0.15* to create varied conditions that encouraged strong generalization.

1. **Noise Factors:** The noise factor was varied across different preprocessing strategies to test the robustness of the CAE model. A proper balance was sought between the degree of augmentation and the inclusion of noise to simulate real-world conditions observed in the nighttime footage.
2. **Data Preparation:** Images were resized to the targeted dimensions and converted to tensor formats suitable for model input. This conversion is vital for enabling the network to process the images in a format it understands.
3. **Impact Assessment:** The chosen preprocessing strategies were evaluated for their impact on model performance, specifically how each approach affected the CAE’s ability to handle variations and noise in the nighttime fridge camera footage.

**Modeling Approach:**

**Model Choice: Convolutional Autoencoder (CAE)**

For this project, we employed the Convolutional Autoencoder (CAE) technique to address the challenges posed by noisy, low-resolution nighttime footage used to identify the ice cream thief. The choice of CAE was primarily driven by its strengths in image processing and feature extraction, which are critical for our task. Below are the key reasons and architectural details that influenced our decision:

1. **Image Compression and Reconstruction:** The CAE effectively compresses images into a lower-dimensional latent space while preserving essential features. This ability is vital for reconstructing clearer facial characteristics from the low-quality surveillance footage.
2. **Spatial Feature Preservation:** The convolutional structure of the CAE enables it to maintain the spatial hierarchy of images, which is crucial for facial recognition. Unlike fully connected networks, convolutions specifically target local patterns and structures necessary for distinguishing faces from blurry backgrounds.
3. **Robustness to Noise:** Given the low-light conditions of the footage, the CAE’s robustness in handling noisy data is particularly beneficial. The convolutional layers enhance the model's capacity to filter out irrelevant noise and focus on significant facial features during the denoising process.
4. **Layered Architecture:** Our CAE model comprises four convolutional layers followed by batch normalization, Leaky ReLU activation, maximum pooling, and dropout. This layered approach enables the model to progressively downsample the input images from 128x128 pixels to an 8x8 feature map, capturing essential facial representations with minimal information loss.
5. **Decoder Functionality:** The decoder mirrors the encoder’s structure with transposed convolutional layers to reconstruct the images back to their original dimensions. The use of sigmoid activation in the output layer ensures that pixel values are properly scaled to reflect realistic image qualities.
6. **Latent Space Encoding:** The latent space consists of a compact representation ranging from 128 to 384 dimensions, effectively encoding the necessary features of the faces for accurate identification. This compactness reduces overfitting risks, especially relevant given the small dataset of only 30 images per individual.

The following CAE architectures were then created:

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Member** | **Model Type** | **Latent Size** | **Activation** | **Normalization** | **Dropout** |
| Prince | Classic CAE | 128 | ReLU | None | None |
| Debjani | Improved CAE | 256 | LeakyReLU | BatchNorm2d | * 1. 0.3 |
| Kathryn | Improved CAE | 128 | LeakyReLU | BatchNorm2d | 0.0-0.4 |

**Hyperparameter Sweeps and Tuning Strategy**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Member** | **Search Method** | **Latent Vector** | **Learning Rates** | **Noise Levels** | **Batch Sizes** |
| Prince | Random | 64-256 | 1e-3, 1e-4 | 0.1-0.3 | 32, 64 |
| Debjani | Random | 96-288 | 1e-2, 5e-4, 1e-4 | 0.1-0.2 | 16, 32 |
| Kathryn | Bayesian | 64-256 | 1e-3, 5e-4, 1e-4 | 0.05-0.2 | 16, 64 |

To optimize the performance of our Convolutional Autoencoder (CAE) models, we conducted extensive hyperparameter sweeps utilizing Weights & Biases. Each team member employed tailored search methods to explore variations in key hyperparameters, such as latent vector sizes, learning rates, noise levels, and batch sizes. For instance, while Prince employed a random search approach with a learning rate range of 1e-3 to 1e-4, Debjani explored a broader latent space and utilized more aggressive learning rates. Kathryn implemented a Bayesian search for efficiency, targeting a learning rate range of 1e-3 to 1e-4. All sweeps ran for 40 epochs with the aim of minimizing validation mean squared error (MSE). This careful tuning process enabled us to identify configurations that yielded the best test performance, ultimately enhancing the model's capability to denoise and reconstruct the blurry footage effectively.

**Results and Interpretation:**

**Evaluation Metrics**

To evaluate the performance of our CAE models, we used the following metrics:

* **Mean Squared Error (MSE)** – Measures the average squared difference between predicted and actual values.
* **Mean Absolute Error (MAE)** – Captures the average magnitude of the errors.
* **R² Score (Coefficient of Determination)** – Represents how well the model explains the variability in the target variable.

A lower MSE and MAE, coupled with a higher R² score (closer to 1), indicates better model performance.

**Summary of Team Results:**

After extensive hyperparameter sweeps, we identified each team member's best configuration based on test MSE performance.

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Member** | **Run Name** | **Latent Dim** | **Learning Rate** | **Dropout** | **Noise** | **Test MSE** | **Test MAE** | **R2 Score** |
| Prince | CAE\_crrux15s | 128 | 0.001 | None | 0.3 | 0.02432 | 0.1116 | 0.608 |
| Debjani | ImprovedCAE\_whhootdi | 288 | 0.0005 | 0.1 | 0.1 | 0.01878 | 0.10314 | 0.690 |
| Kathryn | ImprovedCAE\_zufl6ccx | 128 | 0.0005 | 0 | 0.2 | **0.0173** | **0.09690** | **0.695** |

**Model Configuration Efficacy:** Kathryn's model outperformed the others, achieving the lowest test MSE (0.0173) and the highest R² score (0.695), despite using a moderate latent vector size of 128. This suggests that the model's architecture and hyperparameters were aptly tuned to excel in reconstructing noisy images, highlighting the importance of optimizing model complexity alongside performance.

**Learning Rate Impact**: The results indicated that moderate learning rates, particularly 0.0005, yielded superior performance in reconstruction quality compared to more aggressive rates. Both Debjani and Kathryn's configurations, which utilized this learning rate, demonstrated lower error metrics. This finding suggests that slower learning rates may help the model converge more steadily and effectively in the context of our dataset.

**Noise-Dropout Balance**: The success of Kathryn's model, which incorporated a noise level of 0.2 without dropout, emphasizes the significance of balancing noise injection with regularization techniques. Excessive dropout can hinder the model's learning process by restricting the amount of informative data available during training. This balance is crucial for models that need to learn from noisy data, as seen in our nighttime footage.

**Latent Space Exploration**: Debjani's model, which explored a larger latent space (288), performed well but did not surpass Kathryn's results. This suggests that while expanding the latent dimensions can capture more complex features, an overly large latent space might lead to diminished returns in performance, potentially causing overfitting when the dataset is limited.

**Comparative Analysis of Architectures**: The differences between the classic CAE used by Prince and the improved CAE architectures employed by Debjani and Kathryn highlighted the advantages of incorporating advanced activation functions like LeakyReLU and Batch Normalization. These enhancements contributed to better learning stability and performance under challenging conditions like noise and low lighting.

**Conclusion**

In conclusion, our exploration of using Convolutional Autoencoders (CAEs) to identify the ice cream bandit among roommates proved to be a successful endeavor. Through diligent preprocessing, careful architectural choices, and systematic hyperparameter tuning, we achieved notable improvements in reconstructing blurry nighttime footage. Kathryn's model demonstrated the effectiveness of balancing latent space size, learning rate, and noise levels, resulting in the highest reconstruction quality as evaluated by our metrics.

Our findings highlight the importance of model optimization techniques in enhancing performance under challenging conditions, paving the way for accurate identification of individuals in low-visibility scenarios. Ultimately, this project not only accomplished its goal of unveiling the elusive ice cream thief but also contributed valuable insights into the application of deep learning techniques in image processing and recognition tasks. The journey showcases the power of advanced machine learning methods in resolving real-world challenges, reinforcing the potential for continued research and application in this fascinating domain.

Guardians of The Galaxy