Research

Scalable Real-Time Emotion Recognition using EfficientNetV2 and Resolution Scaling

Base Model: EfficientNetV2

EfficientNetV2 was chosen as the base model as it has a lower computational cost and performs well on a variety of different datasets. In addition, the model has low inference times and therefore is a great option for a real time solution. [1]

Key Techniques

a. Resolution Scaling

- Adjust input resolution to improve accuracy.
- More flexibility makes it easier to use on different hardware.

b. Data Augmentation

- Prior to training images were rotated, flipped, etc: to simulate real world conditions.

c. Training Setup

- Used pre-trained image-net weights to accelerate training.
- Optimized with the Adam optimizer and a dynamic learning rate.
- Models trained for 120 epochs on the KDEF dataset.

Key Points

- **Real-time Execution:** Real time inference time was determined to be 40 ms.
- **Scalability:** Resolution scaling maintained performance across hardware with varying computational capabilities. It was successfully tested on an Intel-I5 processor.

Real-Time Emotional Analysis from A Live Webcam Using Deep Learning

Base Models: MTCNN and VGG-16

MTCNN was used for face detection while VGG-16 was used for facial emotion recognition classification. [2]

Key Techniques

a. Face Detection and Alignment

- Utilized MTCNN to accurately detect and align faces in live webcam feeds.
- Ensured consistent face positioning to improve classification accuracy.

b. Feature Extraction and Classification

- Employed VGG-16 for extracting deep features from facial images.
- Applied Transfer Learning to fine-tune the pre-trained VGG-16 model on the FER2013 dataset.

c. Real-Time Implementation

- Integrated OpenCV for capturing and processing live video streams.
- Achieved real-time emotion recognition by optimizing the processing pipeline.

Key Points

- **High Training Accuracy:** Achieved 97.23% accuracy on the training set, demonstrating effective learning of facial emotion patterns.
- **Real-Time Performance:** Successfully implemented a system capable of processing live webcam feeds and displaying emotion classifications in real-time.
- **Hybrid Model Efficiency:** Combining MTCNN and VGG-16 provided a balanced trade-off between speed and accuracy, suitable for real-world applications.
- **Applicability:** Potential applications include patient monitoring, security surveillance, and e-learning environments.

Datasets

FER2013 Dataset

Based on our findings, we chose to work with the FER-2013 dataset due to its terms of service and availability. The FER2013 dataset consists of grayscale images of faces, sized at 48x48 px. Faces are categorized into one of 7 discrete emotional states:

- 0 = Angry
- 1 = Disgust
- 2 = Fear

- 3 = Happy
- 4 = Sad
- 5 = Surprise
- 6 = Neutral

The dataset is divided into two main subsets:

- **Training Set:** 28,709 examples
- **Public Test Set:** 3,589 examples

Other Notable FER Datasets

The FER-2013 dataset is readily available online and the dataset is relatively small making it an ideal option for small lightweight models.

1. CK+ (Extended Cohn-Kanade):

- **Description:** Contains both posed and spontaneous facial expressions with detailed action units
- **Differences from FER2013:** Better option for dynamic emotional analysis in comparison to the static images available in FER 2013.

2. JAFFE (Japanese Female Facial Expression):

- **Description:** Comprises 213 images of Japanese female subjects displaying seven emotions.
- **Differences from FER2013:** Dataset lack data variety and is suitable for specific problems.

3. AffectNet:

- **Description:** Large dataset with 1 million samples that are labelled both on the discrete and valence emotion scales.
- **Differences from FER2013:** Data is annotated more richly with more detail

4. RAF-DB (Real-world Affective Faces Database):

- **Description:** 30,000 images collected from the net that are labelled according to the 7 basic discrete emotions.
- **Differences from FER2013:** Dataset simulates real world conditions such as different lighting and backgrounds making it a harder but more accurate benchmark.

Considerations on Dataset Size and Image Resolution

Model Size and Dataset Size

- **Larger Datasets:**
  - Can support larger and more complex models with several layers and parameters.
  - Larger dataset means a more varied training set and therefore a better generalized model.
  - Requires more computation and training time.
- **Smaller Datasets:**
  - Works with smaller models.
  - Can regain performance through techniques like transfer learning and data augmentation.
  - Good for mobile solutions and when computational resources are limited.

Reasons for Choosing Smaller Datasets

- **Resource Constraints:** Used a Google Cloud instance with a T4 GPU (16Gb VRAM) for training and a M3 pro to test inference.
- **Faster Experimentation:** Smaller datasets allow for quicker training and iteration during the development and tuning of models.
- **Availability and Terms of Service**: FER-2013 dataset is readily available for download on Kaggle hub and is relatively small.

Impact of Image Resolution on Model Performance and Data Size

- **Higher Pixel Size (Resolution):**
  - Provides more detailed information and can therefore capture more abstract patterns.
  - Increases the amount of data per image, resulting in larger dataset sizes and higher computational and memory requirements.
  - Increasing resolution means we need more parameters in the input layers to caputre the data in each pixel and therefore more reousrces.
- **Lower Pixel Size (Resolution):**
  - Reduces computational and memory requirements, enabling faster training and inference.
  - May not capture more complex relationships due to lack of data.
  - Helps in scenarios where bandwidth or storage is limited, making it easier to manage and process data.

References

1. O. Ghadami, A. Rezvanian, and S. Shakuri, "Scalable Real-time Emotion Recognition using EfficientNetV2 and Resolution Scaling," 2024 10th International Conference on Web Research (ICWR), Tehran, Iran, 2024, pp. 1-7, doi: 10.1109/ICWR61162.2024.10533360.

2. C. A. Kumar and K. Anitha Sheela, "Real-Time Emotional Analysis from A Live Webcam Using Deep Learning," 2022 3rd International Conference for Emerging Technology (INCET), Belgaum, India, 2022, pp. 1-5, doi: 10.1109/INCET54531.2022.9824894.