Unveiling Faces: Decoding Emotions with AI

By Srinivas Annam, Dec 17th, 2023

Problem Definition

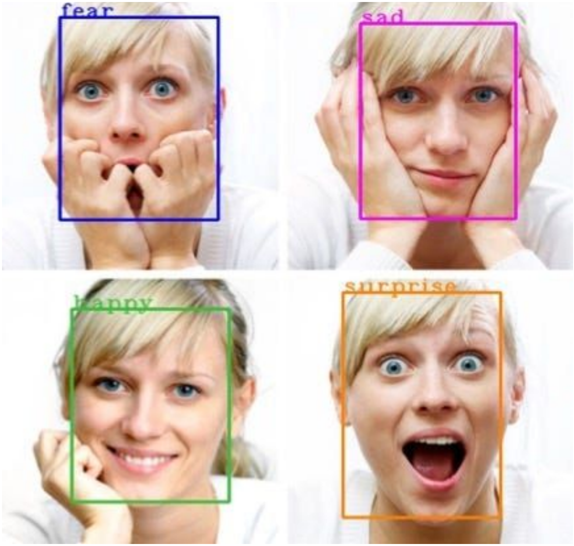
Difficulty in accurately interpreting human emotions can lead to misunderstandings, communication breakdowns, and even poor decision-making.

Emotions are the unspoken language of human interaction. This project explores the power of AI to decipher them from facial expressions.

Problem To Solve

Predict the dominant emotion from a human face image using deep learning models built from scratch or using transfer learning techniques, enabling applications like improved customer service or personalized healthcare.

- Predict the dominant emotion from a human face image.
- Explore the effectiveness of transfer learning for emotion recognition.
- Ensure models are compatible with existing compute resources.
- Achieve accuracy levels necessary for a successful real-world application.

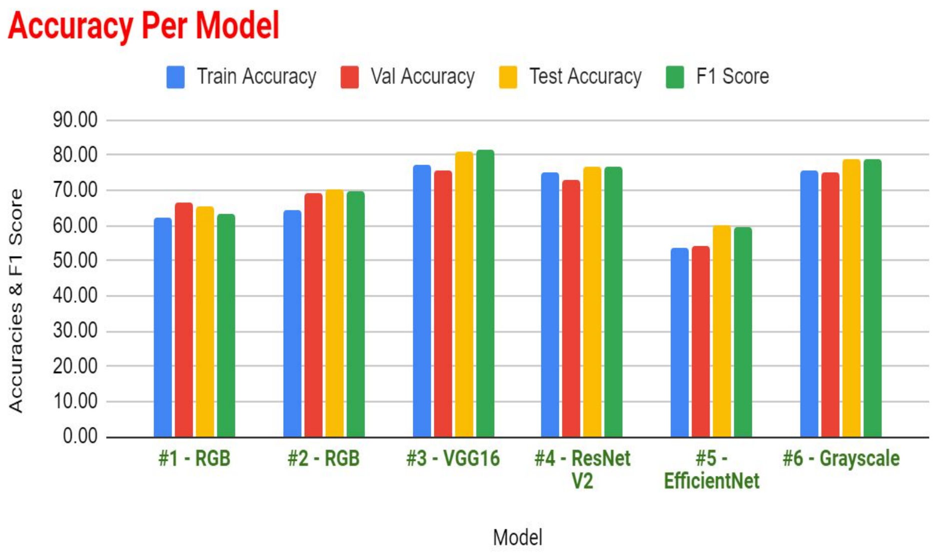


Solution Approach

Model choice: Train from scratch vs. transfer learning (VGG16, ResNet, etc.)

Data Augmentation: Preprocess / rescale images correctly for each transfer learning model.

Two-step Training: 1. Short run for feature extraction, 2. Longer run for emotion recognition fine-tuning.



Model Selection

Model	Train Accuracy	Val Accuracy	F1 Score	Run Time (per epoch)
#1 - 3-block RGB	62.06	66.41	63.16	28s
#2 - 4-block RGB	64.17	69.10	69.72	29s
#3 - VGG16	**77.14**	**75.57**	**81.30**	**28s**
#4 - ResNet V2	75.05	73.10	76.56	52s
#5 - EfficientNet	53.40	54.09	59.61	29s
#6 - 5-block Gray	75.76	75.35	78.73	24s

Transfer learning proved to be more effective compared to training CNN models from scratch! And the right model (VGG16) is not necessarily expensive to run.

Transfer Learning

What is Transfer Learning? Why is it a useful consideration?

- **Leverage existing knowledge:** Reuse pre-trained skills like seasoned athletes learning new sports.
- **Less data, better results:** Learn faster with less data, saving time and resources.
- **Faster learning curve:** Skip basic training and focus on mastering new tricks.
- **Big data advantage:** Leverage vast knowledge from pre-trained models.
- **Adapt to win:** Fine-tune skills for specific tasks, just like athletes adjust for different sports.

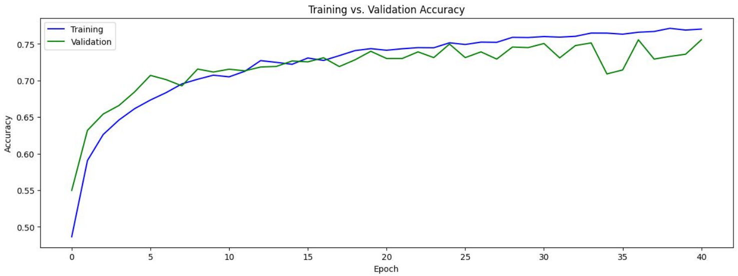
Final Model Solution

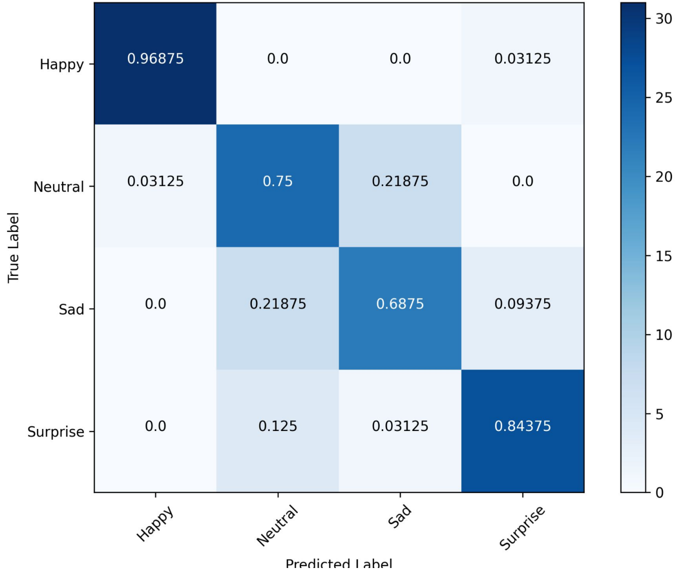
A transfer learning model with the base coming from VGG16.

Augment the data by using pre-built preprocessing function of the model.

Make an initial shorter run of the model for feature extraction.

Make a longer run for fine tuning and finalizing the model.





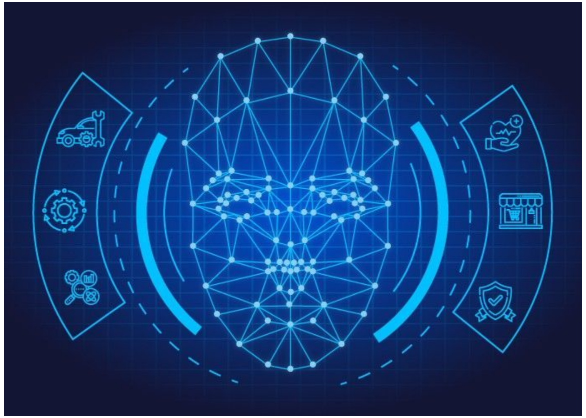
Proposed Business Solution

Develop a system to accurately detect and classify human emotions: *Happy*, *Neutral*, *Sad*, and *Surprised*.

Optimized for efficiency, our model delivers *accurate* results with minimal hardware requirements.

The system will be built using deep learning and artificial intelligence techniques, and it will be capable of recognizing emotions in *real-time*.

Built for versatility, our solution integrates seamlessly into diverse applications like video conferencing, customer service, and security.



Executing Business Solution

Acquire a larger dataset of facial emotions with a higher resolution.

Invest in sufficient GPU resources to enable efficient training and execution of the model for real-world applications.

Train the model on the dataset of facial images with labeled emotions using the best GPU that money can buy.

Deploy the final model to the selected production environment where it can be used to recognize emotions in real-time.

Executive Summary

Key Takeaway: Leverage the VGG16 architecture with transfer learning, facial emotion recognition can be done accurately and efficiently.

Our Solution: A CNN trained on a varied facial expression dataset using transfer learning to achieve accuracy and efficiency.

Benefits: Valuable for a variety of applications, including video conferencing and security.

Next Steps: Implement the proposed solution. Explore additional applications and use cases. Collaborate with stakeholders to ensure ethical and responsible implementation of the solution.

Risks & Challenges

Data Bias: The model's performance might be impacted by the inherent biases present in the training data. Addressing such biases requires thoughtful data selection and augmentation strategies.

Overfitting: Overfitting to the training dataset poses a risk of poor performance on previously unseen data. Techniques like data augmentation and regularization are essential to mitigate overfitting.

Privacy Concerns: Collecting and storing facial images raise privacy concerns that must be addressed. Implementing robust data security measures and following ethical data collection practices are critical.

Explainability and Interpretability: Understanding the model's decision-making process is vital for building trust and ensuring fairness. Research and advancements in explainable AI techniques are needed to address this challenge.

Thank you!

🛑STOP🛑

NOTE: The following slides are not part of the original presentation but rather used as supplementary data as relevant questions come up.

Choice of Hardware

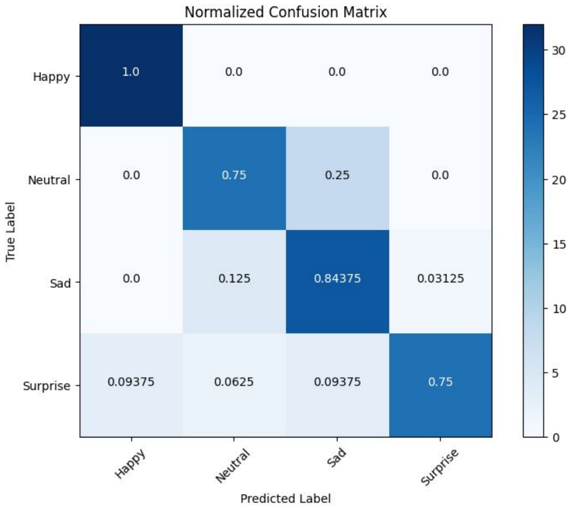
How does the choice of hardware (GPU) can have a significant impact on the performance of a facial emotion recognition model?

GPUs are much faster than CPUs for processing large amounts of data, which is necessary for facial emotion recognition models.

GPUs with more cores and memory will be able to process data more *quickly* and *efficiently*.

The type of GPU (e.g., NVIDIA, AMD) can also affect performance and accuracy of the models built across the same number of epochs.

Overall, using a GPU will result in a significant speedup in the training and inference of facial emotion recognition models.



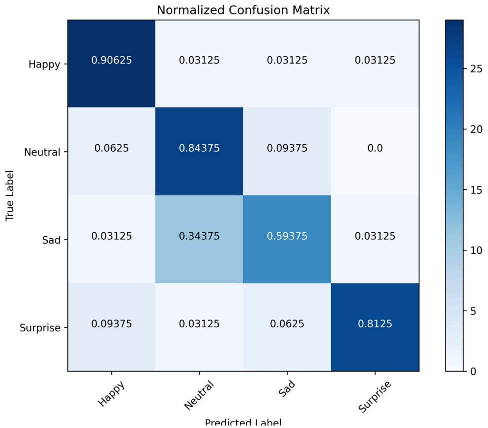
On A100 GPU, VGG16 produced 83.59% vs. 81.30 on a V100 GPU.

Second Best Model

How did training a model from scratch work out? How far is it from the selected model?

VGG16 outperformed Model #6 by a significant margin on the same hardware. The VGG16 model achieved an F1 score of 81.30% on the test set, while Model #6 achieved an F1 score of 78.73%. This difference is likely due to the fact that VGG16 is a more complex model with more parameters, which allows it to learn more detailed features from the data. Additionally, VGG16 was trained on a larger dataset than Model #6, which gave it more opportunity to learn from a wider variety of facial expressions.

I had an earlier more complex model that has since been simplified, but it is very clear that a more complex model that is run on the right GPU can possibly compete or outperform VGG16.



Model #6 performance on V100

RGB vs Grayscale

How did training a model from scratch work out? How far is it from the selected model?

Model	Train Accuracy	Val Accuracy	Test Accuracy
#1 - RGB	62.06	66.41	65.63
#1 - Grayscale	60.51	65.14	67.18
#2 - RGB	64.17	69.10	70.31
#2 - Grayscale	64.84	70.14	71.88

Both Model #1 and #2 showed better performance on test accuracy with the grayscale models