## Deep Learning Assignment 1
## Name: Riyyan Ahmad Bhatti
## Roll no: 22i-2069
## Section: DS-D

# Network Details:

## Dataset

- **Size (after filtering):** 3999 images
- **Labels:** Expression classification (7 classes), Valence & Arousal regression (continuous).
- **Input Shape:** (224, 224, 3)
- **Outputs:**
    - class_output: categorical softmax
    - reg_output: linear (valence, arousal)

## Architecture

Two CNN baselines were tested:

- **VGG16**: Standard 16-layer network pretrained on ImageNet. Top layers replaced with custom dense heads for multi-task learning (classification + regression).
- **ResNet50**: 50-layer residual network with transfer learning from ImageNet, followed by multi-task heads.

## Training Settings

- Optimizer: Adam
- Batch Size: 32
- Epochs: 20
- Learning Rate: 1e-4
- Loss: Weighted combination of categorical cross-entropy (classification) and MSE (regression).
- Hardware: CPU-based training (high training time per epoch ~18–42 min for VGG16, ~7–14 min for ResNet50).

## Rationale for Baseline

VGG16 was chosen as a classical CNN baseline to establish a performance benchmark. ResNet50 was added for comparison, as residual connections often improve convergence and generalization.

# Dataset Splits

- Training: 80% (≈3200 images)
- Validation: 20% (≈800 images)
- Test set results were not reported; validation metrics were used for comparison.

# Transfer Learning

- **VGG16:** Initialized with ImageNet weights but trained on full dataset without freezing layers.
- **ResNet50:** Initialized with ImageNet weights. Transfer learning helped faster convergence and higher accuracy.

# Training Graphs

- **VGG16:** Training accuracy started at 13% and reached ~32% after 20 epochs. Loss steadily decreased. Validation accuracy peaked at ~30%.

- **ResNet50:** Training accuracy increased rapidly, reaching >98% by epoch 20. Validation accuracy plateaued around 34–37%.

# Performance Measures & Discussion (15 pts)

### VGG16

- **Classification:** Accuracy = 28.25%, F1 = 0.225, AUC = 0.74, PR-AUC = 0.305, Kappa = 0.18.
- **Regression:** RMSE = 0.401, CORR = 0.619, SAGR = 0.72, CCC = 0.503.

### ResNet50

- **Classification:** Accuracy = 40.75%, F1 = 0.397, AUC = 0.773, PR-AUC = 0.412, Kappa = 0.323.
- **Regression:** RMSE = 0.369, CORR = 0.687, SAGR = 0.756, CCC = 0.637.

### Continuous Domain Metrics – Rationale

- **RMSE:** Measures magnitude of regression error. Useful for minimizing deviations in predicted valence/arousal.

- **CORR:** Captures linear relationship between predicted and ground-truth values.
- **SAGR (Sign Agreement):** Checks if predicted emotional polarity (positive/negative) matches ground truth. Important in real-world affective systems.
- **CCC (Concordance Correlation Coefficient):** Combines accuracy and correlation; best suited for continuous affect prediction "in the wild".

**Most Suitable Metric:** CCC, as it balances correlation and error, ensuring reliability for real-world affective computing systems.

## Performance Comparison of CNN Architectures

| Model | Accuracy | F1-macro | AUC | RMSE | CORR | SAGR | CCC | Training Time/Epoch |
|---|---|---|---|---|---|---|---|---|
| VGG16 | 28.25% | 0.225 | 0.74 | 0.401 | 0.619 | 0.72 | 0.503 | ~18–42 min |
| ResNet50 | 40.75% | 0.397 | 0.773 | 0.369 | 0.687 | 0.756 | 0.637 | ~7–14 min |

**Observation:**

- ResNet50 significantly outperformed VGG16 on both classification and regression tasks.
- ResNet50 trained faster per epoch due to optimized architecture and residual connections.
- VGG16 served as a weak baseline, while ResNet50 showed clear benefits of modern architectures.