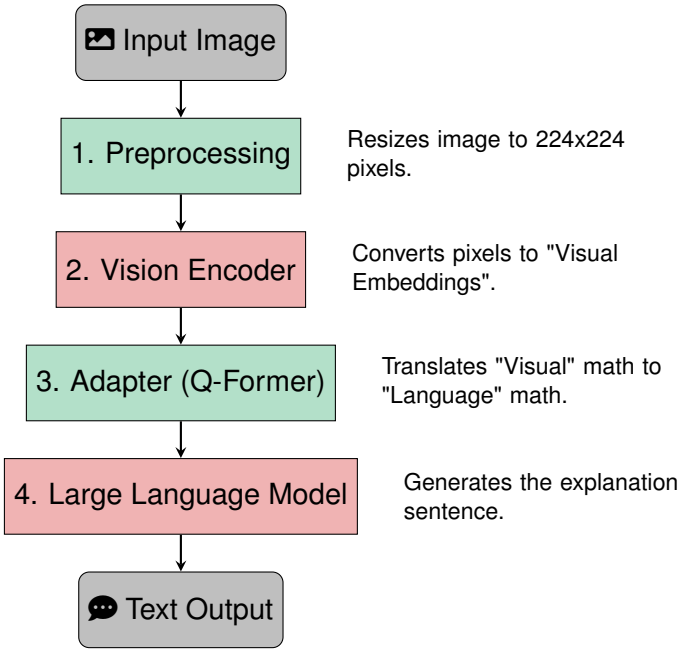
Meow AI: The Technical Pipeline

From Pixels to Understanding

Meow AI Team

This document explains exactly what happens inside the "Black Box" of our AI.



Stage 1: The Input & Preprocessing

What happens: The user uploads a photo (e.g., a selfie).

- **Face Detection:** We crop the image to focus only on the face (removing background noise).
- **Normalization:** We resize the image to exactly `224x224` pixels because the AI model expects this exact shape.
- **Tensor Conversion:** The image is converted from a standard JPG file into a PyTorch Tensor (a matrix of numbers).

Tools: OpenCV, PyTorch Transforms.

Stage 2: The Vision Encoder (The Eye)

What happens: The "Blind" LLM needs to "see" the image.

- We pass the tensor into a **Vision Transformer (ViT)** (like EVA-CLIP).
- **Output:** It doesn't output "Happy" or "Sad" yet. It outputs **Embeddings** (a long list of numbers representing shapes, eyebrows, mouth curves).

Note: We usually *freeze* this part during training to save time.

Stage 3: The Adapter & LLM (The Brain)

What happens: This is where the magic (and our fine-tuning) happens.

- **The Adapter (Q-Former):** It takes the visual embeddings from Stage 2 and reshapes them so they look like "words" to the LLM.
- **The Prompt:** We combine the image info with a text prompt: *"Question: What emotion is this person feeling? Answer:"*
- **The LLM (e.g., LLaMA/Vicuna):** It takes the prompt + image info and auto-completes the sentence.

Our Job: We use **LoRA** to train the connection between the Adapter and the LLM so it learns our specific 14 emotions.

Stage 4: The Output & XAI

What happens: The model generates text.

- **Raw Text:** "The person is happily surprised because..."
- **Parsing:** We look for keywords (e.g., "surprised") to assign the official category Label (0-13).
- **XAI (Explainability):** We run **Grad-CAM** on the Vision Encoder to generate a heatmap, showing which pixels (eyes, mouth) were most important for the decision.

Final Result: Shown to the user on the Streamlit App.