Real or Fake

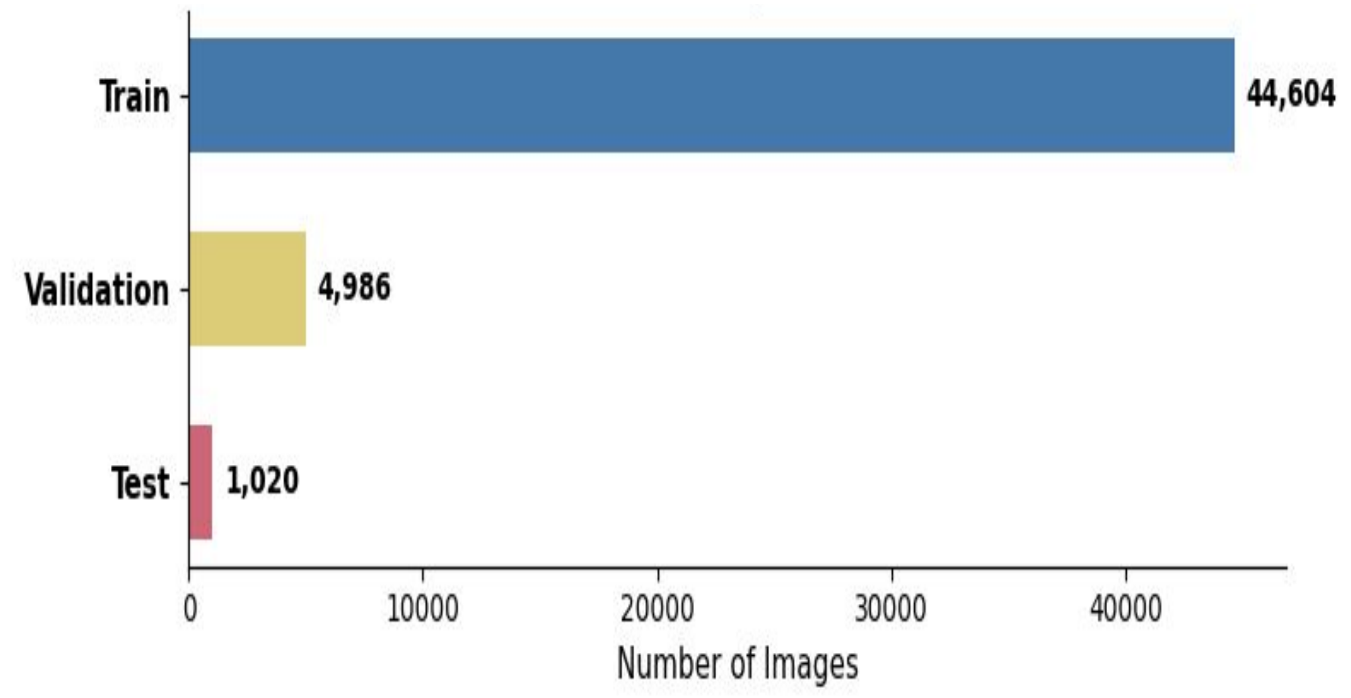
Luke Tu, Renxuan Yao, Qianyun Gong
CSCI 1470 Deep Learning Final Project

Introduction

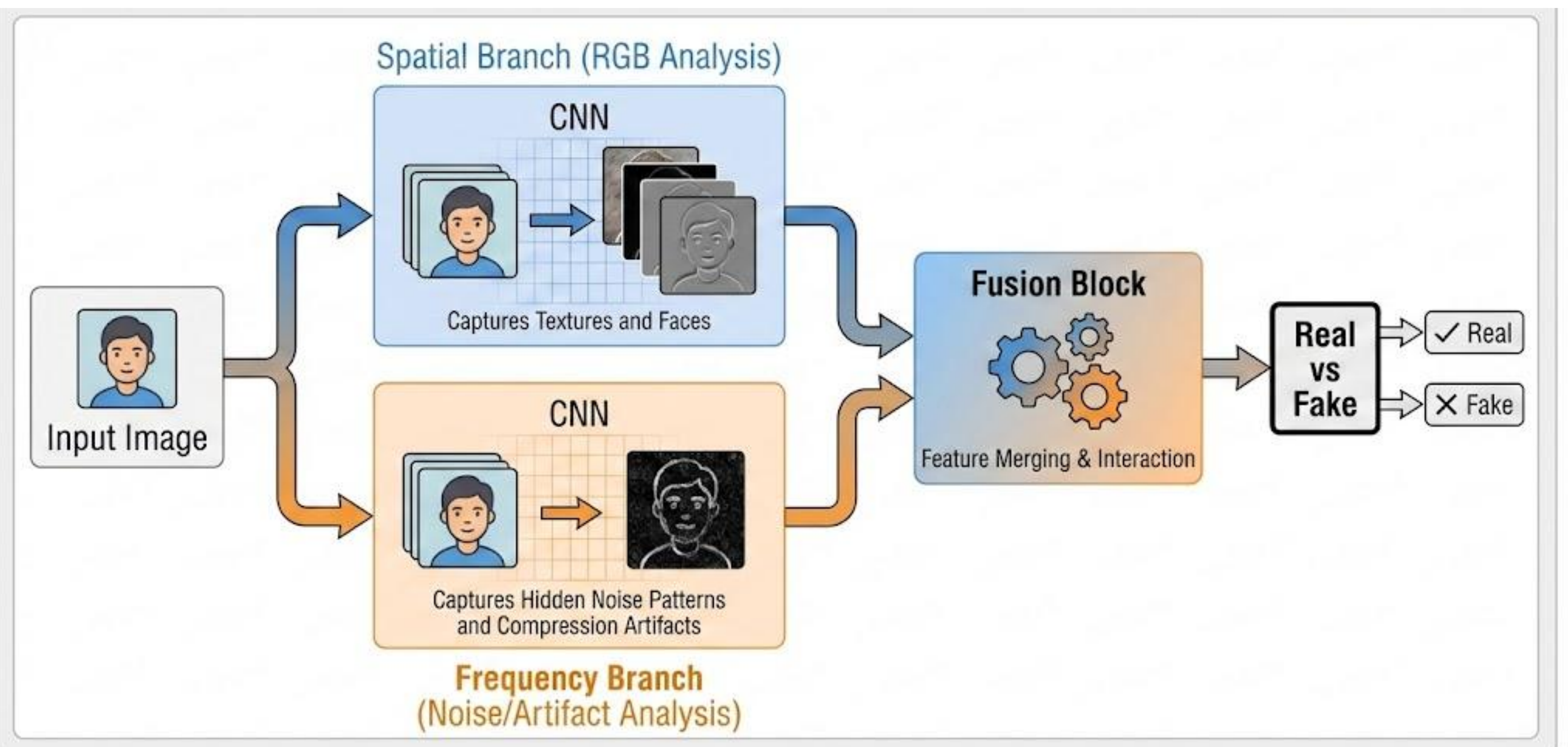
The Problem: The proliferation of AI images and content has grown exponentially in the last few years, as a result of the recent AI boom, making it increasingly difficult for people to distinguish between what's real and what's not.
Why It Matters: This image generation technology brings all sorts of potential problems, such as for misinformation, deep fakes, fraud, and copyright infringement. Platforms, regulators, and enterprises need scalable, automated tools to answer a seemingly simple question: "Is this image real or AI-generated?" Relying on human moderators or manual inspection is too slow, too subjective, and not scalable.

Methodology

Model Architecture: We built a Dual-Branch Neural Network designed to detect AI images not just by how they look (Spatial), but by the hidden mathematical fingerprints they leave behind (Frequency). Then we created a fusion classifier.
Implementation: Our model uses a dual-branch architecture: a Spatial Branch, built on a ResNet-18 backbone, learns visual features such as edges, textures, and RGB patterns. In parallel, a Frequency Branch processes the log-magnitude DFT to capture spectral fingerprints of AI images, including periodic artifacts, overly smooth frequency bands, and missing sensor noise. Their feature vectors are then combined in a Fusion Module, where a small MLP with dropout and batch normalization produces the final classification, achieving higher accuracy than either branch alone. We train the system using binary cross-entropy with the Adam optimizer (lr = 1e-4), apply augmentations like flips, crops, and JPEG compression for robustness, and implement everything in PyTorch with GPU acceleration.



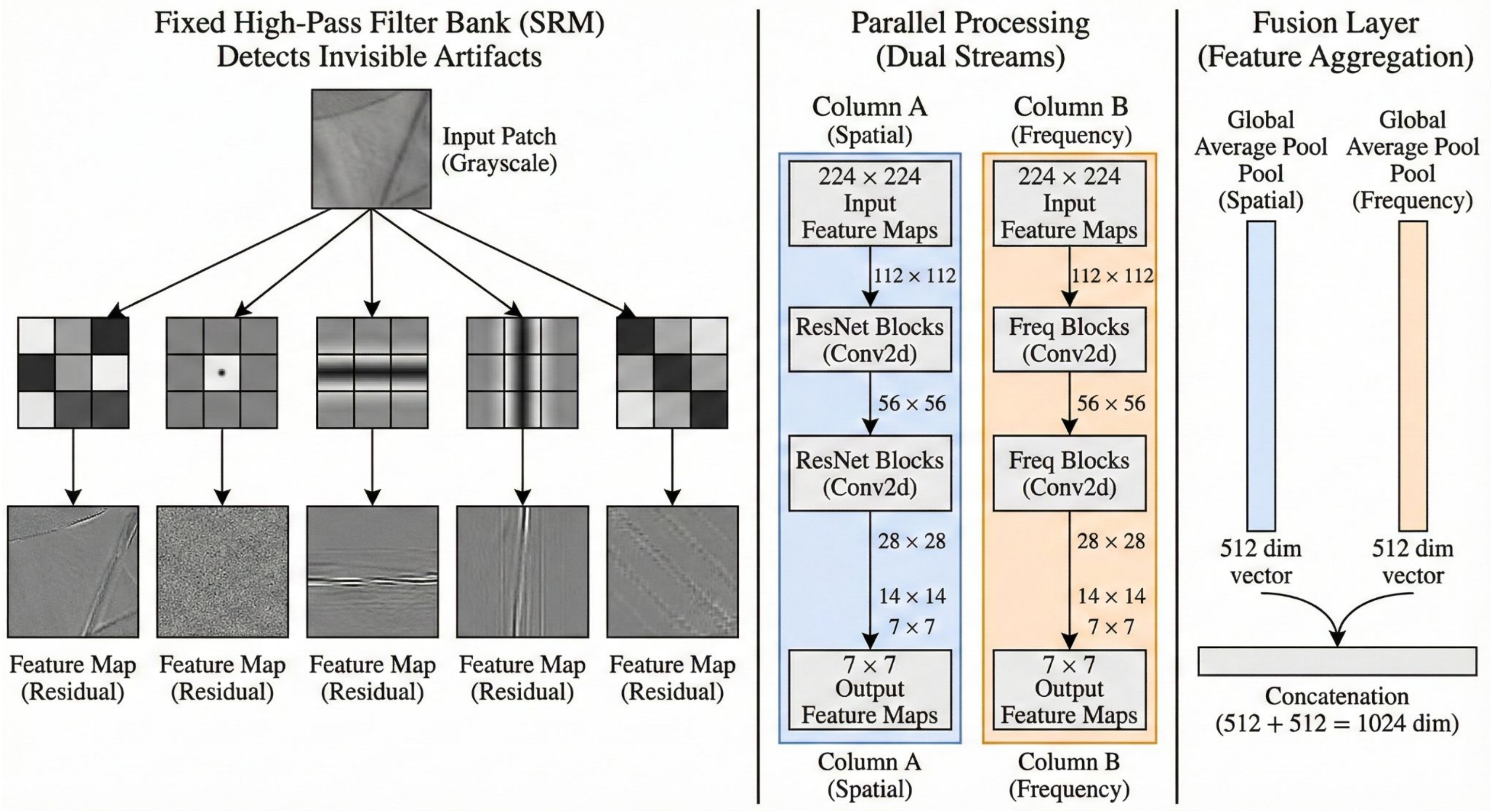
Data Sample Counts


Spatial Branch (RGB Analysis)
CNN
Captures Textures and Faces
Input Image
CNN
Captures Hidden Noise Patterns and Compression Artifacts
Frequency Branch (Noise/Artifact Analysis)
Fusion Block
Feature Merging & Interaction
Real vs Fake
✓ Real
✕ Fake

High-Level Overview: Dual-Stream Deep Learning Architecture

Discussion

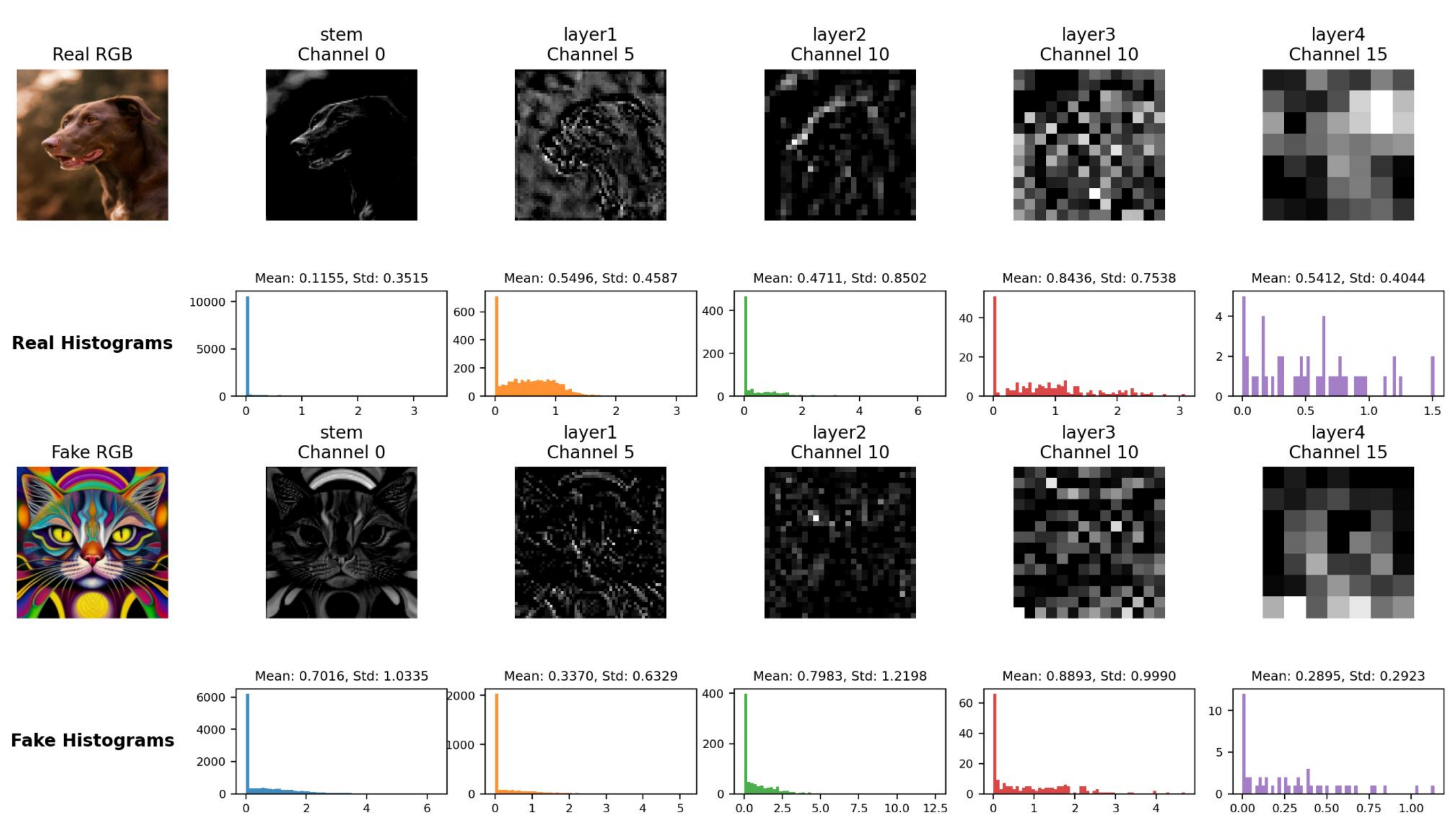
We learned that dynamic, on-the-fly preprocessing is essential for scalability and that proper FFT normalization is required for the Frequency branch to train stably. The current system is limited by sensitivity to JPEG compression and potential bias toward specific generators. Future work includes developing the full fusion model to combine spatial and frequency cues, expanding the dataset to newer diffusion models and compression settings, and deploying a lightweight interface for real-time image verification.



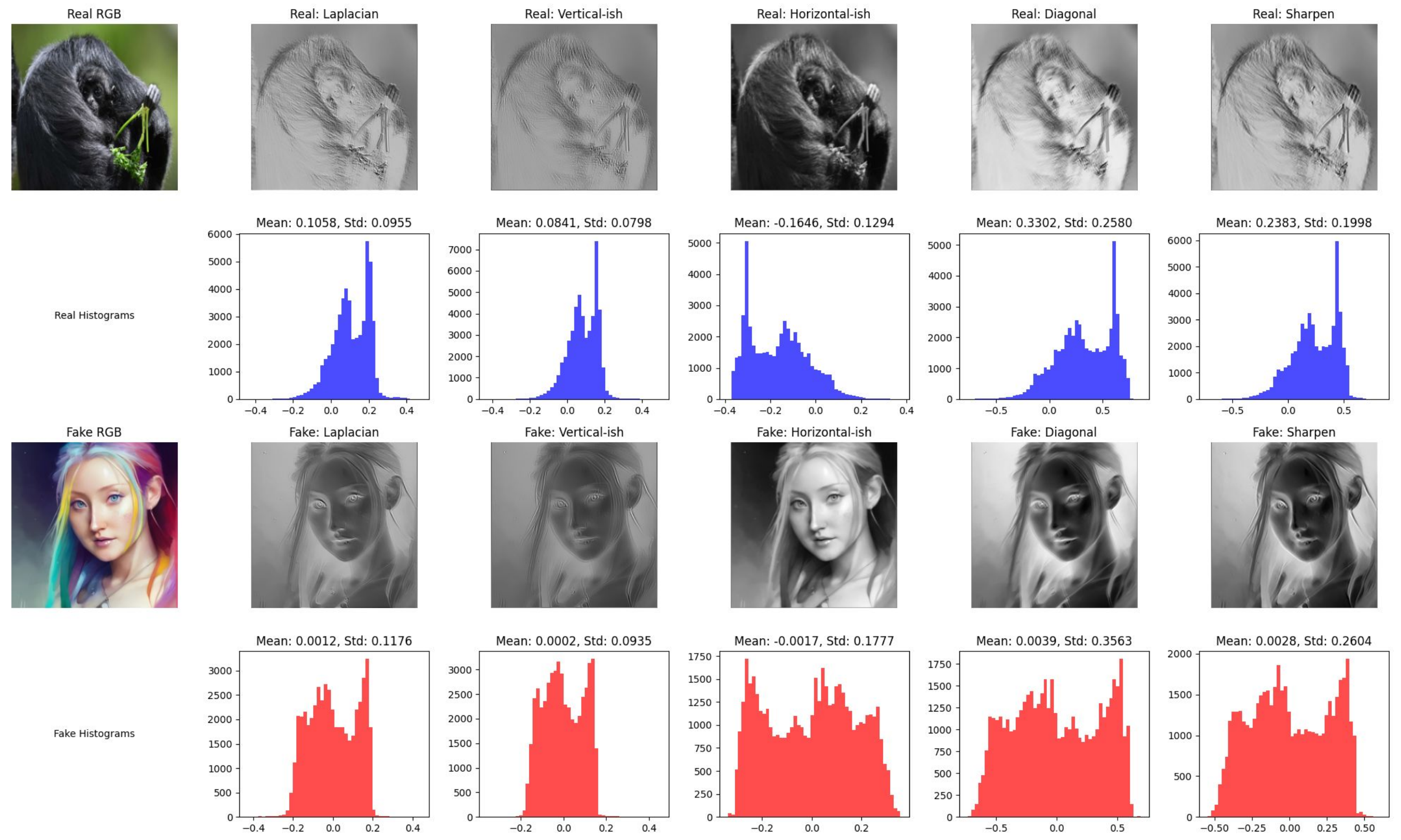
Detailed Architecture: The Feature Extraction & Fusion Mechanism

Dataset

Our dataset combines multiple open-source sources from Kaggle and Hugging Face, covering natural scenes, faces, objects, and artwork, with a balanced mix of real camera images and synthetic outputs from modern diffusion models. The task is a simple binary classification, Real versus Fake, but the key engineering choice was to use **dynamic loading**, where each image is resized, normalized, and FFT-processed on the fly. This avoids storing multiple preprocessed variants, prevents disk overflow, and keeps the entire pipeline scalable to large datasets.



Real and Fake Image Comparison: Spatial Features

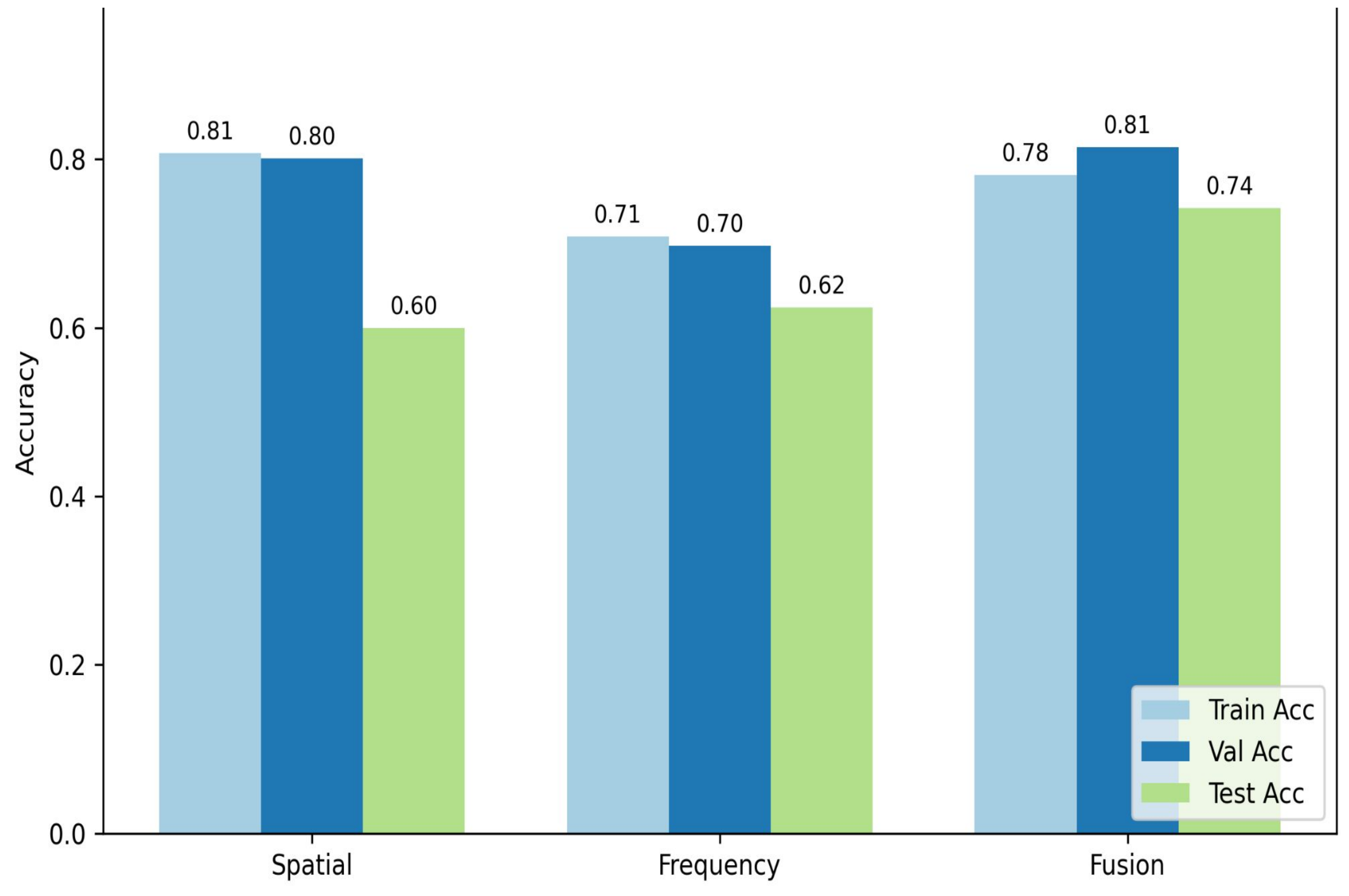


Real and Fake Image Comparison: Frequency Features

Results

	Train		Validation		Test (Unseen)	
Model	Loss	Acc	Loss	Acc	Loss	**Acc**
Spatial (RGB)	0.4448	80.7%	0.4502	80.1%	0.7306	60.0%
Frequency (SRM)	0.5709	70.9%	0.5792	69.8%	0.6534	62.5%
Fusion	**0.4976**	**78.1%**	**0.4484**	**81.5%**	**0.5354**	**74.2%**

Table: Model Accuracy



Histogram: Model Accuracy

Our dual-stream Fusion model achieved a test accuracy of 74.23%, significantly outperforming both the individual Spatial (60.00%) and Frequency (62.45%) baselines. While the single-branch models suffered from overfitting—performing well on training data but dropping sharply on unseen test data—the fusion approach successfully integrated texture and noise features to generalize better.