Benchmarks Directory

This directory contains image quality benchmarking tools for Smart Content Studio's multi-model AI image generation system.

📁 Contents

- `benchmark_image_quality.py` - Main benchmarking script with CLIP and FID evaluation
- `benchmark_requirements.txt` - Python dependencies for benchmarking
- `BENCHMARKING.md` - Comprehensive guide with usage instructions and methodology
- `benchmark_results/` - Generated benchmark results and test images

🚀 Quick Start

From the project root directory:

```
# Install dependencies
pip install -r benchmarks/benchmark_requirements.txt

# Run quick benchmark (4 prompts, ~2 mins)
python benchmarks/benchmark_image_quality.py --quick

# Run full benchmark (12 prompts, ~10 mins)
python benchmarks/benchmark_image_quality.py
```

📊 What It Tests

- **CLIP Score**: Semantic similarity between prompts and generated images
- **FID Score**: Image quality compared to MS-COCO real-world images
- **Multi-provider comparison**: Pollinations, Gemini Imagen, Grok (when available)
- **Quality tier evaluation**: Fast, Balanced, High, Ultra modes

📖 Full Documentation

See `BENCHMARKING.md` for:

- Detailed installation instructions
- Interpretation of metrics
- Advanced usage options
- Troubleshooting guide
- Citation information

⚠ Prerequisites

1. **Backend must be running**: Ensure FastAPI server is active at `http://localhost:8000`

2. **API keys configured**: At least `GEMINI_API_KEY` in `backend/.env`
3. **Python 3.10+**: Required for PyTorch and CLIP
4. **Sufficient disk space**: ~2GB for CLIP models and MS-COCO stats (for FID)

💡 Results Interpretation

CLIP Score (higher is better):

- 0.20-0.25: Acceptable
- 0.25-0.30: Good
- 0.30+: Excellent (industry-leading)

FID Score (lower is better):

- <20: High quality, photorealistic
- 20-30: Good quality
- 30-50: Acceptable quality
- > 50: Poor quality

🔗 Related Documentation

- Main README - Project overview
- Backend README - API documentation
- Benchmark Summary - Latest results (if available)