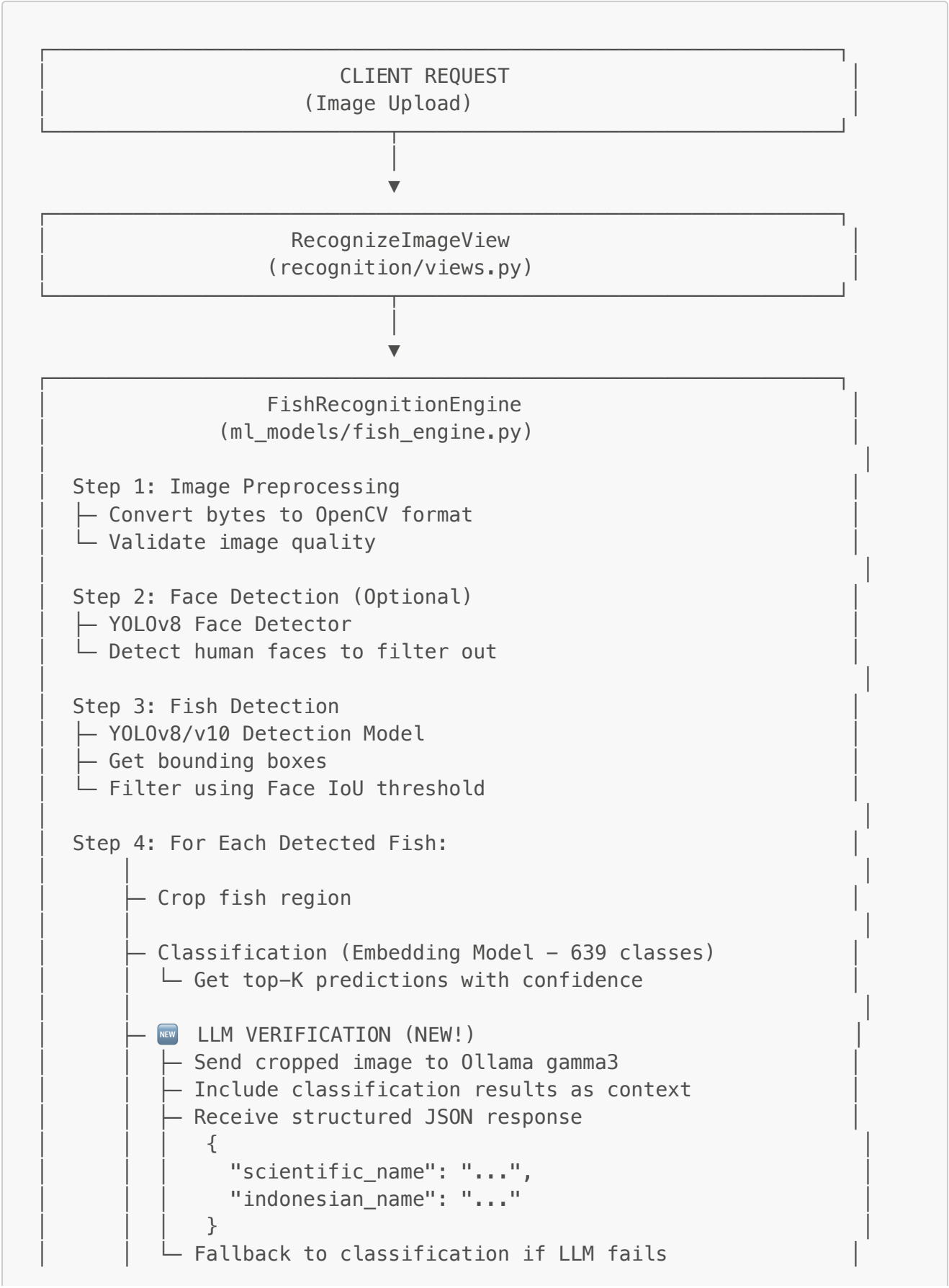
Fish Recognition Flow dengan LLM Enhancement

Architecture Overview



```
│      │                                                                  │
│      └─ Segmentation (Optional)                                         │
│         └─ Generate polygon mask                                        │
│                                                                         │
└─────────────────────────────────────────────────────────────────────────┘
                                  │
                                  ▼
┌─────────────────────────────────────────────────────────────────────────┐
│                    RESPONSE TO CLIENT                                   │
│                                                                         │
│  {                                                                      │
│    "success": true,                                                     │
│    "fish_detections": [                                                 │
│      {                                                                  │
│        "id": 0,                                                         │
│        "bbox": [x1, y1, x2, y2],                                        │
│        "confidence": 0.95,                                              │
│        "classification": [                                              │
│          {                                                              │
│            "label": "Oreochromis mossambicus",                          │
│            "confidence": 0.85                                           │
│          }                                                              │
│        ],                                                               │
│        "llm_verification": {  // 🆕 NEW                                 │
│          "scientific_name": "Oreochromis mossambicus",                  │
│          "indonesian_name": "Ikan Mujair",                              │
│          "processing_time": 2.5                                         │
│        },                                                               │
│        "segmentation": {...}                                            │
│      }                                                                  │
│    ]                                                                    │
│  }                                                                      │
└─────────────────────────────────────────────────────────────────────────┘
```

LLM Service Architecture

```
┌─────────────────────────────────────────────────────────────────────────┐
│              FishRecognitionEngine                                      │
│                                                                         │
│  process_image()                                                        │
│    │                                                                    │
│    ├─ Detection & Classification                                        │
│    │                                                                    │
│    └─ if llm_enabled:                                                   │
│         └─ llm_service.verify_classification()                          │
│              │                                                          │
└──────────────┼──────────────────────────────────────────────────────────┘
               │
               ▼
┌─────────────────────────────────────────────────────────────────────────┐
```

```
│           OllamaLLMService                                  │
│     (services/ollama_llm_service.py)                        │
│                                                             │
│  verify_classification(image, detection_info)               │
│     │                                                       │
│     ├─ Convert image to base64                              │
│     ├─ Build system prompt                                  │
│     ├─ Build user prompt with context                       │
│     │   - Top classifications                               │
│     │   - Detection confidence                              │
│     │                                                       │
│     └─ Send to Ollama API                                   │
│          │                                                  │
└──────────┼──────────────────────────────────────────────────┘
           │
           ▼

┌─────────────────────────────────────────────────────────────┐
│         Ollama API (gamma3 model)                           │
│      https://ollama.hellodigi.id                            │
│                                                             │
│  POST /api/generate                                         │
│  {                                                          │
│    "model": "gamma3",                                       │
│    "prompt": "Identifikasi ikan ini...",                    │
│    "system": "Anda adalah ahli identifikasi ikan...",       │
│    "images": ["base64_encoded_image"],                      │
│    "format": "json"                                         │
│  }                                                          │
│                                                             │
│  Response:                                                  │
│  {                                                          │
│    "response": "{                                           │
│      \"scientific_name\": \"...\",                          │
│      \"indonesian_name\": \"...\"                           │
│    }"                                                       │
│  }                                                          │
└─────────────────────────────────────────────────────────────┘
```

System Prompts

System Prompt (untuk Ollama gamma3)

```
Anda adalah ahli identifikasi ikan yang sangat akurat.
Tugas Anda adalah mengidentifikasi spesies ikan dari gambar yang
diberikan.

PENTING:
- Berikan HANYA nama ilmiah (scientific name) dan nama Indonesia
- Jika ada beberapa ikan, identifikasi yang paling dominan/jelas
- Jika tidak yakin atau bukan ikan, kembalikan null
```

```
- Format output HARUS JSON: {"scientific_name": "...",
"indonesian_name": "..."}
- JANGAN tambahkan penjelasan atau text tambahan
- Jika tidak dapat diidentifikasi: "Unknown" dan "Tidak dikenal"
```

User Prompt (contoh)

```
Identifikasi ikan dalam gambar ini.

Model klasifikasi memprediksi:
- Oreochromis mossambicus (85.23%)
- Oreochromis niloticus (10.45%)
- Clarias gariepinus (2.31%)

Confidence deteksi: 95.00%

Berikan identifikasi Anda dalam format JSON yang diminta.
```

Configuration & Management

Environment Variables

```
OLLAMA_URL=https://ollama.hellodigi.id
OLLAMA_MODEL=gamma3
LLM_ENABLED=True
LLM_TIMEOUT=30
```

Management Endpoints

1. **GET /api/recognition/config/llm/**
    - Check LLM status
    - View configuration
    - Health check
2. **POST /api/recognition/config/llm/**

    ```
    {"enabled": true/false}
    ```

    - Enable/disable LLM dynamically
    - No restart required
3. **GET /api/recognition/health/**

- Overall system health
- Includes LLM status

4. **GET /api/recognition/stats/**
    - Performance metrics
    - Includes LLM processing times

Error Handling Flow

```
LLM Verification Attempt
    │
    ├─ Success ──────────────────────► Return LLM result
    │
    ├─ Timeout (>30s) ───────────────► Return null + log warning
    │
    ├─ Service Unavailable ──────────► Return null + log error
    │
    ├─ Invalid JSON Response ────────► Return null + log error
    │
    └─ Network Error ────────────────► Return null + log error

In all error cases:
✓ Classification results still returned
✓ API request completes successfully
✓ LLM is enhancement, not requirement
```

Performance Considerations

Processing Times (Estimates)

- Detection: ~50-100ms
- Classification: ~100-200ms
- LLM Verification: ~2000-5000ms ⚠
- Segmentation: ~200-300ms
- **Total: ~2500-5600ms per image**

Optimization Strategies

1. **Parallel Processing**: Run LLM verification in parallel for batch
2. **Selective LLM**: Only verify low-confidence classifications
3. **Caching**: Cache LLM results for similar images
4. **Timeout Control**: Adjust based on use case

Use Cases

High Accuracy Required

```
# Enable LLM for critical identifications
POST /api/recognition/config/llm/
{"enabled": true}
```

Real-time Speed Required

```
# Disable LLM for faster processing
POST /api/recognition/config/llm/
{"enabled": false}
```

Hybrid Approach

- Use LLM only when classification confidence < 0.7
- Implement in client or add threshold configuration