

Collection Execution Plan

🎯 Objective

Collect **1,000-1,500** high-quality basketball shooting images for RoboFlow model training.



Timeline: 5-7 Days

Day 1: Setup & Initial Collection

Morning (2 hours)

- Run `setup_api_keys.py` (15 min)
- Collect from Pixabay (200 images, 15 min)
- Collect from Pexels (200 images, 15 min)
- Collect from Unsplash (200 images, 15 min)
- Run Vision AI filter (40 min)
- **Expected output:** 400 filtered images (60-70% acceptance rate)

Afternoon (2 hours)

- User approval session 1 (150-200 images)
- **Expected output:** 150 approved images

Day 1 Total: 150 approved images

Day 2: More Collection + Approval

Morning (1.5 hours)

- Collect from Pixabay (new query, 200 images, 15 min)
- Collect from Pexels (new query, 200 images, 15 min)
- Run Vision AI filter (30 min)
- **Expected output:** 250 filtered images

Afternoon (2 hours)

- User approval session 2 (200 images)
- **Expected output:** 180 approved images

Day 2 Total: 330 approved images (cumulative)

Day 3: Additional Sources

Morning (2 hours)

- Collect from Unsplash (new queries, 200 images, 20 min)
- Run Vision AI filter (40 min)
- User approval session 3 (150 images, 1 hour)
- **Expected output:** 140 approved images

Afternoon (2 hours)

- User approval session 4 (200 images)
- **Expected output:** 180 approved images

Day 3 Total: 650 approved images (cumulative)

Day 4: Gap Analysis & Targeted Collection**Morning (1 hour)**

- Run `progress_tracker.py`
- Analyze approved images:
 - Check angle diversity (side, front, 45°)
 - Check player diversity (male, female, youth)
 - Check shooting phase distribution (prep, release, follow-through)
- Identify gaps

Afternoon (3 hours)

- Targeted collection for gaps:
 - Specific search queries for missing categories
 - Additional API calls (if rate limits allow)
 - YouTube video frame extraction (if needed)
 - Vision AI filter (30 min)
- User approval of targeted images (2 hours, 150 images)
- **Expected output:** 140 approved images

Day 4 Total: 790 approved images (cumulative)

Day 5: Final Push to 1,000**Morning (2 hours)**

- Collect final batch (varied sources, 300 images, 30 min)
- Vision AI filter (30 min)
- User approval session 5 (150 images, 1 hour)
- **Expected output:** 130 approved images

Afternoon (2 hours)

- User approval session 6 (remaining images, 100 images)
- Final review and quality check
- **Expected output:** 90 approved images

Day 5 Total: 1,010 approved images (cumulative)

Day 6-7: Buffer & Optimization (Optional)**If target not reached:**

- Continue collection from alternative sources
- Lower Vision AI threshold slightly (e.g., 45 instead of 50)
- Re-review rejected images for edge cases

If target reached:

- Organize images by category
 - Prepare for RoboFlow upload
 - Create annotation guidelines
-

**Expected Metrics****Collection Performance**

Metric	Expected	Actual
Raw images collected	1,500-2,000	
Vision AI acceptance rate	60-70%	
Filtered images	900-1,400	
User approval rate	75-85%	
Final approved images	1,000-1,200	

Time Breakdown

Activity	Time per 100 images	Total Time (1,000 images)
Collection	5-10 min	50-100 min
Vision AI filtering	15-20 min	150-200 min
User approval	15-20 min	150-200 min
Total		5.5-8.3 hours

**Optimization Strategies****Speed Up Collection****1. Run multiple collectors in parallel:**

```
```bash
Terminal 1
python multi_source_collector.py -source pixabay

Terminal 2
python multi_source_collector.py -source pexels
```

```
Terminal 3
python multi_source_collector.py -source unsplash
```

```

1. Use rate limit efficiently:

- Collect during off-peak hours
- Spread collection across multiple days
- Use different API keys if available

Speed Up Vision AI Filtering

1. Batch processing:

- Process 50 images at a time
- Use parallel processing (if API allows)

2. Smart filtering:

- Skip images with low file size (<50 KB)
- Pre-filter by image dimensions (min 800x600)

Speed Up User Approval

1. Use keyboard shortcuts:

- A (approve) and R (reject) are much faster than clicking
- Shift+Click for bulk selection

2. Review in focused sessions:

- 30-60 minutes per session
- Take 5-10 minute breaks
- Review 100-150 images per session

3. Sort by quality:

- Review highest-scoring images first
- Reject lowest-scoring images in bulk

⚠ Risk Mitigation

Risk 1: API Rate Limits

Mitigation:

- Spread collection across multiple days
- Use multiple API keys (if allowed)
- Implement exponential backoff
- Monitor rate limit headers

Risk 2: Low Vision AI Acceptance Rate

Mitigation:

- Adjust search queries for better results
- Lower quality threshold (e.g., 45 instead of 50)
- Manually review some rejected images
- Use more specific keywords ("free throw" vs "basketball")

Risk 3: User Approval Fatigue

Mitigation:

- Review in short sessions (30-60 min)
- Take breaks between sessions
- Use keyboard shortcuts for speed
- Gamify the process (track approval speed)

Risk 4: Image Quality Issues

Mitigation:

- Increase minimum resolution requirements
 - Filter out low-quality sources
 - Use professional stock photo APIs
 - Manually curate a “golden set” of 100 images
-



Success Criteria

Quantity

- ✓ **1,000+ approved images** (minimum for MVP)
- ⌚ **1,500+ approved images** (target for production)

Quality

- ✓ **Basketball visible:** 100%
- ✓ **Full body visible:** >90%
- ✓ **Clear shooting form:** >85%
- ✓ **Average Vision AI score:** >70

Diversity

- ✓ **Angles:** Side (40%), Front (30%), 45° (30%)
 - ✓ **Players:** Male (60%), Female (30%), Youth (10%)
 - ✓ **Phases:** Preparation (30%), Release (40%), Follow-through (30%)
 - ✓ **Sources:** Distributed across 3+ sources
-



Progress Tracking

Daily Check-in Questions

1. How many images collected today?
2. What is the Vision AI acceptance rate?
3. How many images approved by user?
4. What is the cumulative total?
5. Are we on track for the target?
6. Any bottlenecks or issues?

Weekly Review

- Review progress dashboard

- Analyze quality metrics
 - Identify weak categories
 - Adjust collection strategy
-



Completion Checklist

- [] 1,000+ images approved by user
 - [] All images organized by category
 - [] Metadata files complete and accurate
 - [] Progress report generated
 - [] Quality metrics meet targets
 - [] Diversity requirements met
 - [] Images ready for RoboFlow upload
 - [] Annotation guidelines prepared
-

Last Updated: December 13, 2025

Status: Ready for execution

Estimated Completion: Day 5-7