

FLUX AI Selection Pipeline

Phase 1 – Build the Training Dataset

Folder Structure

```
PHILLY_IN_FLUX/ ├── market-st/ | ├── originals/ | └── selected/ | ├──  
germantown-ave/ | ├── originals/ | └── selected/ | ├── frankford-ave/ |  
├── originals/ | └── selected/ | ├── washington-ave/ | ├── originals/ |  
└── selected/ | ├── ridge-ave/ | ├── originals/ | └── selected/ | ├──  
passyunk-ave/ | ├── originals/ | └── selected/ | ├── lancaster-ave/ | ├──  
originals/ | └── selected/ | ├── walnut-st/ | ├── originals/ | └── selected/  
| ├── girard-ave/ | ├── originals/ | └── selected/
```

Goal

For every project:

- originals = every photograph shot (~1000)
- selected = photographs accepted into archive (~150)

Nothing else.

This is your ground truth.

Phase 2 – Generate Labels

Create a script that scans both folders.

Output:

```
filename,label IMG_0001.JPG,0 IMG_0002.JPG,1 IMG_0003.JPG,0
```

Where:

- 1 = archive selection
- 0 = rejected

Goal:

10,000 originals 1,500 selected

Phase 3 — Enrich Metadata

For every image:

Extract:

- EXIF
- GPS
- Timestamp
- Camera settings

Store:

```
{ "filename": "...", "selected": true, "gps": "...", "timestamp": "...", "camera": "...",  
  "metadata": {...} }
```

Phase 4 — Generate AI Vision Descriptions

Run every image through a vision model.

Generate tags such as:

rowhouse storefront church window doorway vacant lot crosswalk pedestrian fence
graffiti utility pole

Store alongside metadata.

This creates future archive search capability.

Phase 5 — Train FLUX Selector

Input:

Image + Metadata

Output:

Archive Probability

Example:

IMG_1234.JPG → 0.98 IMG_1235.JPG → 0.91 IMG_1236.JPG → 0.03

The model learns your archive threshold.

Not your best photograph.

Your keep/reject decision.

Phase 6 — Automated Ingest

Future workflow:

Walk Street ↓ Shoot 1000 Photos ↓ Insert SD Card ↓ Import to FLUX ↓ Metadata Extraction ↓ Vision Analysis ↓ Selection Model Runs ↓ Top 150 Chosen ↓ Project Created ↓ Map Generated ↓ Statistics Generated ↓ Archive Ready

Human review:

150 images ↓ Approve ↓ Publish

No more manually reviewing 1000 photographs.

Immediate Next Action

Do not train AI yet.

Do not build the ingest system yet.

First build:

10 Projects ↓ Originals Folder ↓ Selected Folder ↓ Labels CSV

Once that dataset exists, Claude can build everything else from it.