GenAI for Packaging Layout Interpretation

Introduction

This project set out to create a GenAI-assisted pipeline that could analyze images of luxury pen box packaging and translate them into structured layout representations. The approach combines computer vision techniques with AI to detect components, understand their spatial relationships, and output structured data that could be used by downstream applications.

1. Image Analysis Approach

Technology Selection

I selected **YOLO (You Only Look Once)** for this project after evaluating several object detection approaches. YOLO offers several advantages for packaging layout analysis:

- **Speed**: YOLO processes images in a single pass through the neural network
- **Accuracy**: Recent YOLO versions provide excellent detection performance
- **Adaptability**: Can be fine-tuned on custom datasets with relatively few examples
- **Component Detection**: Well-suited for identifying discrete packaging elements

Specifically, I used YOLOv11n, which balances speed and accuracy with a relatively lightweight architecture (2.58M parameters), making it suitable for real-time applications.

Dataset Creation and Training

I created a custom dataset by annotating images of luxury pen boxes with the following component classes:

- Pen
- Insert (the internal holder)
- Base (bottom part of box)
- Lid (top part of box)
- Outer (external structure)

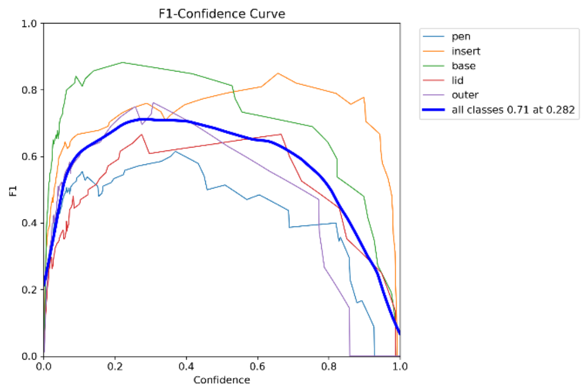
The model was trained for 100 epochs, achieving strong performance:

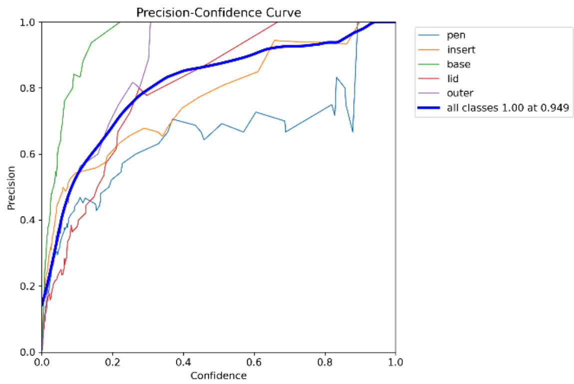
- Overall mAP50 (mean Average Precision at IoU 0.5): 0.76
- Component-specific performance:
  - Insert: 0.874 mAP50
  - Base: 0.913 mAP50
  - Lid: 0.650 mAP50

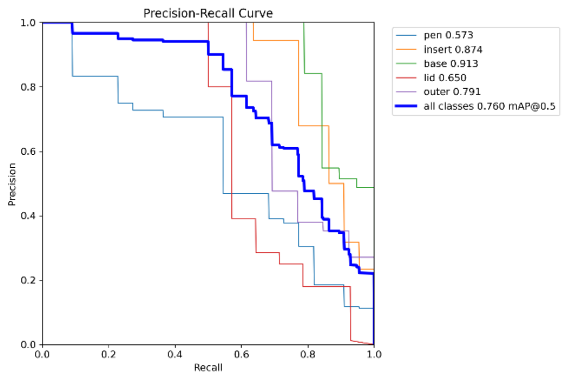
- Outer: 0.791 mAP50
- Pen: 0.573 mAP50

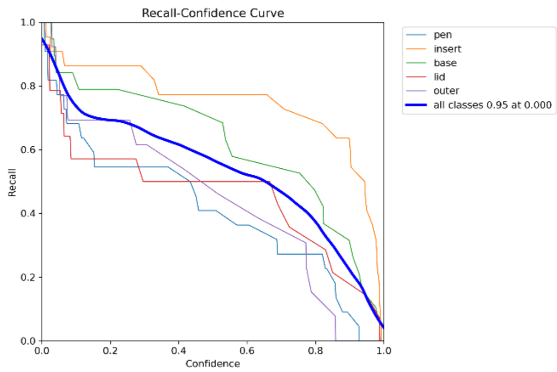
Performance Analysis

The model performed particularly well on structural elements like the base and insert but had lower accuracy for pen detection. This makes sense as pens can vary significantly in appearance, while structural components tend to have more consistent shapes and positions.









Test - Image Results

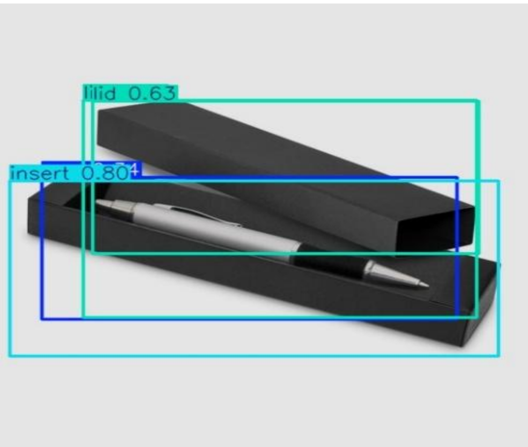
The image shows the output from our trained YOLOv11n model on a luxury pen box. As visible in the visualization, the model successfully detected and classified the key components of the packaging:

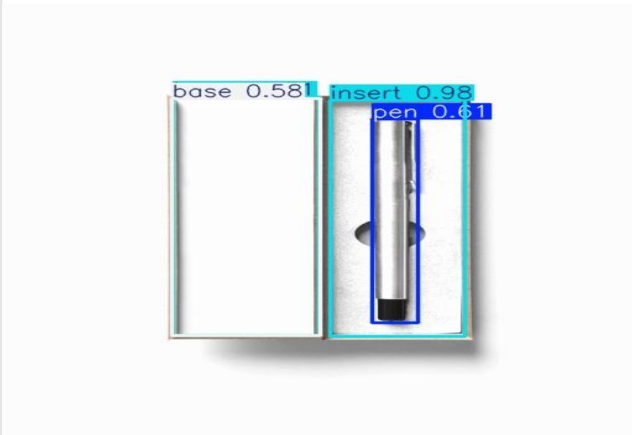
The image shows the output from our trained YOLOv11n model on a luxury pen box. As visible in the visualization, the model successfully detected and classified the key components of the packaging:

The detection results demonstrate the model's ability to identify multiple packaging elements simultaneously, with bounding boxes indicating each component's location and dimensions. The model achieved reasonable confidence scores across all components:

- The lid (0.54 confidence) is correctly identified as the top portion of the red box
- The insert (0.57 confidence) is accurately detected as the black velvet component holding the pen
- The pen itself (0.37 confidence) is identified despite its reflective metallic surface
- The outer structure (0.55 confidence) is properly detected as the packaging boundary

Notably, the model performs better on structural components with consistent shapes (insert, lid, outer) compared to the pen, which aligns with our training metrics where pen detection showed lower mAP scores. This visualization confirms that our object detection approach can effectively decompose a packaging image into its constituent parts, forming the foundation for our structured layout representation. Below are few more samples from testing:





2. Structure Representation

JSON Layout Format

Detection results are transformed into a structured JSON format containing:


```
{
  "insert" : {
    "dimensions" : [
      0 : 388
      1 : 245
    ]
    "type" : "pen_holder"
    "material" : "black velvet"
  }
  "outer" : {
    "dimensions" : [
      0 : 379
      1 : 312
    ]
  }
  "lid" : {
    "dimensions" : [
      0 : 422
      1 : 334
    ]
  }
  "base" : {
    "dimensions" : [
      0 : 387
      1 : 287
    ]
  }
  "pen" : {
    "dimensions" : [
      0 : 320
      1 : 225
    ]
    "relative_position" : "centre"
  }
}
```


Download JSON

Spatial Relationship Analysis

Beyond simple detection, the system calculates relative positioning between components. For example, the pen's position relative to the insert is calculated by comparing their center coordinates and classifying the relationship as "centre," "left corner," or "right corner."

User Interface

I developed a Streamlit application that allows users to:

1. Upload images of pen boxes

2. Process them through the trained model
3. Visualize the detection results
4. View and download the structured JSON representation

3. AI Roadmap for Future Development

Image Preprocessing

AI can be used to enhance image quality, remove background noise, and normalize image color and lighting conditions. This step improves detection accuracy downstream.

Object Detection & Segmentation

Modern computer vision models like YOLOv11 or Detectron2 can detect packaging components such as:

- Outer box (lid and base)
- Internal insert (pen holder)
- Accessories or dividers

These models can also return bounding boxes and confidence scores.

Material and Texture Identification

AI models trained on labeled material data (e.g., PU leather, velvet, paperboard) can classify the visible surfaces of packaging components. Tools like CLIP or vision transformers can be adapted for this.

Dimension & Position Estimation

Based on bounding box coordinates and relative image size, AI can estimate component dimensions and infer relative positions (e.g., "pen is centered inside insert").

Layout Structuring

Post-detection, AI can translate image-based results into structured formats (e.g., JSON or Python dictionaries) that represent packaging layout data — component types, materials, and layout hierarchy.

Generative Design

With ControlNet or Stable Diffusion, GenAI could visualize or even suggest new packaging designs based on an interpreted layout.

Generative Capabilities

The pipeline could be extended with generative components:

- **Layout Generation**: Using diffusion models to suggest design variations
- **3D Model Creation**: Converting detected layouts to CAD-compatible 3D models
- **Design Optimization**: Recommending improvements based on material usage or structural integrity

Required Training Data

To evolve the system, we would need:

- **Diverse Packaging Examples**: Thousands of images covering various styles and configurations
- **Material Samples**: Labeled examples of different packaging materials
- **CAD-Paired Dataset**: Images matched with their corresponding design files
- **Dimension Ground Truth**: Actual measurements paired with visual data

Validation Approach

4. How Would You Validate the Outputs?

Validation ensures reliability and accuracy. It can happen at several levels:

Detection Validation

- Use standard metrics like **IoU (Intersection over Union)**, precision, and recall
- Check model confidence scores

Layout Validation

- Verify if the spatial relationships and component types are logically consistent (e.g., insert should be inside base, not outside)
- Rule-based assertions using layout metadata

Material Validation

- Compare AI-predicted materials with human-labeled textures
- Use confusion matrix to track common misclassifications

Human-in-the-Loop

- Use tools like Streamlit for interactive feedback
- Allow design teams to verify, correct, or override AI outputs

Conclusion

This prototype demonstrates that GenAI, particularly computer vision models like YOLO, can effectively interpret packaging layouts from images. The integration of AI and GenAI into the packaging layout interpretation pipeline holds immense promise — not just for automating manual tasks, but for redefining how packaging is conceptualized, validated, and even designed. This project lays the groundwork by demonstrating how object detection and layout structuring can be achieved using computer vision models like YOLOv8. With further integration of GenAI tools such as ControlNet and GPT-based systems, we can unlock new capabilities like generative design, smart validation, and real-time layout recommendations.

By leveraging structured data, high-quality annotations, and domain-specific rules, we move toward a future where machines don't just see packaging — they understand it.