**Image recognition:**

Models tried:

1. florence-2-base:
   1. {'img': None, 'text': '{\'<MORE\_DETAILED\_CAPTION>\': \'The image is a flowchart that represents the process of creating a product order. It is a visual representation of the steps involved in the process.\\n\\nThe flowchart is divided into four sections. The first section is labeled "Customer", the second section is titled "Place a Product Order", and the third section is labelled "Finish". \\n\\nThere are four steps in the flowchart:\\n\\n1. Sales:\\n2. Confirm if the product is received\\n3. Check the inventory\\n4. Check a credit card\\n5. Check it in stock\\n6. Deliver the order\\n7. Processing the payment\\n8. Finance:\\n9. Finance is the process where the product has been processed.\\n10. Check credit cards\\n11. Check your credit card.\\n12. Check cash in the stock\\n13. Check payment.\\n14. Check for the purchase\\n15. Check on the credit card and check on the purchase.\\n16. Check in the order.\\n17. Check off the purchase in stock.\\n18. Check if the purchase is in stock."\\n19. Check whether the purchase will be processed or processed.\'}'}
   2. PROBLEMS here:
      1. Detected 4 sections, but listed only 3
      2. Detected sections columnwise, not rowise for the diagram
      3. Not detected smaller (YES/NO) tags
      4. Cant give input prompt as we need (Restricted to limited input tasks, like “Caption, Detail Caption, Dense region identification, etc.”

But what we need is understanding of image end to end along with its process flow

1. Llaav-13B:
   1. The flow diagram depicts a process for checking the inventory and managing orders. Here is a sequence of events and actions based on the diagram: 1. Check the inventory: The process starts with checking the inventory to confirm if the product is in stock. 2. Confirm if the product is in stock: If the product is in stock, proceed to the next step. If not, the customer will be informed, and the order will be canceled. 3. Place a product order: If the product is in stock, the customer places an order for the product. 4. Confirm the order: The order is confirmed, and the customer is informed. 5. Process the order: The order is processed, and the product is prepared for shipment. 6. Ship the product: The product is shipped to the customer. 7. Receive payment: The customer receives the product and makes a payment, either through a credit card or another payment method. 8. Confirm payment: The payment is confirmed, and the order is considered complete. In summary, the flow diagram illustrates a process that starts with checking the inventory, followed by confirming the order, processing the order, shipping the product, receiving payment, and finally, confirming the payment to complete the order.
   2. Problems:
      1. Gives descriptive enough answers but not reliable enough
      2. Missed the starting and ending point detection
      3. Uses its own understanding of flows, not capturing the flow we have in diagram so is less accurate
      4. Gets stuck in loop at times to give output:
         1. 12. Check the credit card 13. Yes or No 14. If Yes, proceed to the next step 15. Is the card valid? 16. Yes or No 17. If Yes, proceed to the next step 18. Yes or No 19. If Yes, proceed to the next step 20. Yes or No 21. If Yes, proceed to the next step 22. Yes or No
         2. Multiple Yes No flows unnecessarily
2. Google Cloud Vision API:
   1. Good for image detection, not for understanding flow or detecting sequences
3. *Cogvlm:*
   1. The image depicts a flowchart of a customer's journey with a company. The sequence starts with the customer placing a product order. The company then confirms the order receipt. The customer is then prompted to check the product's availability in stock. If the product is in stock, the customer proceeds to check the card's validity. Once the card is valid, the order is processed for delivery. The delivery is then confirmed, and the order is finally completed.
   2. Adv:
      1. Good in understanding the process
   3. Missing smaller details like who processed what part of sequence
4. Blip2-
   1. Tried in simple mode as other was paid one (gave output in 10-20words)
   2. Problem:
      1. Hard to figure in simple mode if model is good or not.
      2. Gives single word answers, cant describe diagrams properly in detail
5. Idefics3-8B-Llama3:
   1. Step 1: The image is a swimlane diagram with both horizontal and vertical orientations. Step 2: There are four main categories of swinlanes: CUSTOMER, SALES, STOCKS, FINANCE. - In the CUSTOMER category, there's an activity "Place a Product Order". - In the SALES category, there are two activities: "Confirm if order is received" and "Cancel the order" - In the STOCKS category, there are three activities: "Check the inventory", "Is the product in stock?" and "Deliver the order" - In the FINANCE category, there are two activities: "Check credit card" and "Processing the payment" Step 3: - From the CUSTOMER swimlane, Place a Product Order leads to Confirm if order is received which then goes into Is the product in stock? If Is the product in stock? is No, it cancels the order; otherwise, Check credit card. - From Check credit card, if the answer is Yes, Processing the payment occurs, while if the answer is no, Deliver the order happens. The flow direction for most connections is from left to right except when going back to cancel the order. Answer: If the customer places an order, check if the order was confirmed by sales. Then verify that the item is available in stocks. If not in stock, you should cancel the order. Otherwise, validate the client’s credit card information. If valid, proceed with processing the payment and delivering the order.
   2. Adv:
      1. Good enough understanding of diagram
      2. Understands sequence of flow
      3. Captures small elements
      4. Open-source
   3. Problems:
      1. Sometimes missed 1 or 2 blocks while understanding image (Still can be tailored using proper prompts I believe)
      2. Takes around 70-80sec to understand 1 image
6. Blip3:
   1. Horizontal swimlane diagram Customer is connected with Place a Product Order which is then connected with Sales. Sales is connected with Confirm if received order which is then connected with Is the product in stock?. If Is the product in stock? is No then Cancel the order and if Is the product in stock? is YES then Check the inventory which is then connected with Finance. If Finance is Is the card valid? then Check credit card which if Check credit card is No then Is the product in stock? and if Check credit card is YES then Processing the payment which is then connected with Deliver the order.
   2. Adv:
      1. Good understanding of process
      2. Descriptive output (better than blip2)
      3. Fast output (4.5sec)
   3. Problems:
      1. Image flow understanding sometimes go left and right. Doesn’t understand at times. If arrow is not pointed in either vertical or horizontal flow, it assumes connection of some blocks.
7. Gemini API Call directly:
   1. 1.5-pro:
      1. Misses multiple blocks
      2. Missed some arrows
      3. Presumes some flow of diagram based on its understanding of words
      4. Took 4.5sec to process 1img
   2. 1.5-flash:
      1. Near perfect results
      2. Outstanding understanding of flow
      3. Took 3.5sec to understand 1img
      4. Captures decision points properly
   3. 1.5-flash-8b:
      1. Very fast - 2.8sec
      2. Good understanding of images
      3. Can’t understand flow diagram entities like what decision point is and what is normal block in swim lane