

**SINDH MADRESSATUL ISLAM UNIVERSITY**   
**KARACHI**

QUIZ NO: **03** 

SUBJECT:    
**(NATURAL LANGUAGE PROCESSING) NLP**

SUMITTED TO SIR:    
**SYED HASSAN ZAIDI** 

SUBMITTED BY:   
Muhammad Saad Naseem

SEMESTER: **8TH SPRING**   
STUDENT ID: **BSE-21F-097**  
DEPARTMENT OF **BS SOFTWARE ENGINEERING**

**Research Overview**

**1.1 Model Architectures (2020–2025)**

* **Early Approaches (2020–2022):**
* **CNN + LSTM/Transformer** (e.g., PlotQA, DVQA): Extracted visual features using CNNs and processed questions with sequential models.
* **BERT-based models**: Used for text understanding but struggled with chart structure.
* **Recent Advancements (2022–2025):**
* **Vision-Language Transformers** (Pix2Struct, MatCha): Unified image-text encoding.
* **Graph-Based Models** (ChartFormer, ChartNet): Explicitly model relationships between chart elements.
* **Diffusion Models** (Emerging): Used for chart synthesis and augmentation.

**1.2 Reasoning Capabilities**

* **Basic QA (2020–2022):**
* Simple retrieval (e.g., "What is the value of X?").
* Limited support for comparisons, trends, or multi-step reasoning.
* **Advanced Reasoning (2023–2025):**
* **Chain-of-Thought (CoT)** (ChartQA, DePlot): Step-by-step reasoning.
* **Numerical & Logical Reasoning** (MatCha, ChartLlama): Math operations and conditional logic.
* **Multi-Hop Reasoning** (ChartNet): Combining data from multiple chart regions.

**1.3 Chart Analysis Techniques**

* **OCR-Centric Methods (Early 2020s):**
* Relied heavily on OCR for text extraction, leading to errors in complex charts.
* **Structure-Aware Parsing (2023–2025):**
* **Visual Decomposition**: Segmenting charts into axes, legends, bars, etc.
* **Semantic Graph Construction**: Representing charts as graphs for relational reasoning.

Shape

**2. Gap Identification**

**2.1 Limitations in Reasoning Complexity**

* **Gap:** Most models struggle with **multi-hop, numerical, and logical reasoning** (e.g., "If trend X continues, what will the value be in 2026?").
* **Evidence:**
* ChartQA (2022) introduced complex questions but relied on rule-based post-processing.
* MatCha (2023) improved numerical reasoning but still lacks robust logical inference.

**2.2 Efficiency Issues**

* **Gap:** Heavy models (e.g., Pix2Struct) are slow for **real-time applications** (e.g., mobile or edge devices).
* **Evidence:**
* Most SOTA models use large transformers (>100M params), limiting deployment.
* Lightweight alternatives (e.g., MobileViT adaptations) are underexplored.

**2.3 Generalization Across Chart Types**

* **Gap:** Models trained on **bar/line charts** fail on **complex charts** (e.g., radar, waterfall, stacked area).
* **Evidence:**
* ChartVLM (2024) showed improvements but still underperforms on rare chart types.
* Most datasets (ChartQA, PlotQA) lack diversity in chart styles.

**2.4 Weak Multimodal Integration**

* **Gap:** Poor handling of **text + visual + structural** data fusion.
* **Evidence:**
* Early models treated OCR and vision features separately.
* ChartFormer (2023) introduced cross-modal attention but still struggles with noisy OCR.

Shape

**3. Recent Advancements (2022–2025)**

|  |  |  |
| --- | --- | --- |
| **Model** | **Key Innovation** | **Limitations** |
| **Pix2Struct** | Unified image-to-text pretraining | Struggles with complex reasoning |
| **MatCha** | Math-aware pretraining | Computationally expensive |
| **ChartFormer** | Graph-based reasoning | Limited to common chart types |
| **ChartNet** | Multi-hop reasoning with GNNs | Requires large training data |
| **ChartLlama** | LLM fine-tuning for QA | Hallucinations in numerical answers |

**Key Innovations:**

* **Graph-Based Reasoning (ChartFormer, ChartNet):** Explicitly models relationships (e.g., bars → legends).
* **Knowledge-Augmented Models (ChartLlama):** Integrate external knowledge (e.g., finance, statistics).
* **Diffusion for Data Augmentation (2024):** Generating synthetic charts for training.

Shape

**4. Future Directions**

**4.1 Enhancing Reasoning Capabilities**

* **Logical & Multi-Hop Reasoning:** Incorporate **symbolic AI** or **neuro-symbolic** methods.
* **Numerical Robustness:** Improve math operations (e.g., interpolation, extrapolation).

**4.2 Improving Efficiency**

* **Lightweight Hybrid Models:** Combine MobileViT with distilled LLMs.
* **On-Device Optimization:** Quantization, pruning, and efficient attention mechanisms.

**4.3 Handling Diverse Chart Types**

* **Unified Chart Parsing:** Train on **synthetic + real** charts of all types.
* **Few-Shot Adaptation:** Use LLMs to generalize to unseen chart styles.

**4.4 Better Multimodal Fusion**

* **Noise-Robust OCR Integration:** Self-correcting OCR with vision-language alignment.
* **Dynamic Graph Attention:** Adjust graph structure based on question context.