Official Documentation: Visual Dialogue Agent Based on Deep Q Learning and Memory Module Network

1. Introduction

The **Visual Dialogue Agent (VDA)** is an AI system that integrates **Computer Vision (CV)** and **Natural Language Processing (NLP)** to engage in meaningful image-based conversations with users. The system utilizes **Deep Q Learning and Memory Module Networks** to improve dialog quality, understand user preferences, and answer both **Relational** and **Non-Relational** questions about images.

2. Features

- **Image Understanding:** Uses **Convolutional Neural Networks (CNNs)** to analyze images.
- **Conversational Ability:** Engages in multi-turn **dialogues** using NLP techniques.
- **Memory-Driven Responses:** Leverages **End-to-End Memory Module Networks** for relational question answering.
- **User Inclination Awareness:** Utilizes **Deep Q Learning Policies** to understand user preferences.
- **Training on Large Datasets:** Trained on **CLEVR** and **VQA datasets** for robust learning.

3. System Architecture

The system consists of:

- **Image Encoder:** CNN-based encoder for feature extraction.
- **Question Processor:** Uses **Recurrent Neural Networks (RNNs)** for language understanding.
- **Memory Module Network:** Stores conversation history for context-aware answers.
- **Deep Q Learning Module:** Optimizes responses based on user engagement.

4. Datasets Used

- **CLEVR Dataset:** 70,000 training images with **699,989** questions.
- **VQA Dataset:** 265,016 images with multiple questions per image.

5. Implementation Details

- **Programming Language:** Python
- **Deep Learning Frameworks:** TensorFlow / PyTorch
- **Model Training:** CNN for image features, RNN for text, and RL for learning user preferences.

6. Results & Performance

- Achieved **94.4% accuracy** on CLEVR dataset.
- Improved **Relational Question Accuracy** from **73.69% to 75.52%**.
- Enhanced **dialog coherence and user engagement** through reinforcement learning.

7. Future Enhancements

- Improve **user personalization** for tailored responses.
- Extend to **real-world datasets** beyond CLEVR/VQA.
- Optimize **dialog efficiency** using advanced RL techniques.