🗓 IPMV2026 - 2026 8th International Conference on Image Processing and Machine Vision

📍 Da Nang, Vietnam   🗓 Jan 10 - Jan 12, 2026

FLAMINGO-VQA: Modular Input-Output Stabilization for Few-Shot Visual Question Answering

Paper ID: 16

UNDER REVIEW

Abstract

Few-shot Visual Question Answering (VQA) remains a challenging multimodal task, requiring robust reasoning over limited annotated examples. Despite advances, state-of-the-art models like Flamingo, which integrates frozen large language models (LLMs) with a visual encoder via gated cross-attention, struggle with generalization due to noisy, unfiltered visual features. We introduce FLAMINGO-VQA, a novel enhancement to the Flamingo pipeline, optimizing few-shot VQA through three non-trainable techniques: Question-Guided Feature Pre-Selection (QGFP), Semantic Few-Shot Selection (SFS), and Self-Consistency Voting (SCV). QGFP uses a frozen CLIP text encoder to filter irrelevant visual patches, improving input quality. SFS leverages CLIP embeddings to curate semantically relevant exemplars, enhancing context. SCV aggregates stochastic predictions for stability. Designed for computational efficiency, FLAMINGO-VQA requires no retraining, preserving compatibility with existing implementations. Evaluated on the VQA v2.0 benchmark, it achieves a 41.0% accuracy (an 10.0 percentage point improvement over the 31.0% baseline), demonstrating enhanced accuracy, robustness to noise, and generalization across question types. This lightweight, modular framework advances few-shot VQA for resource-constrained environments, offering a scalable solution for real-world applications.

Keywords: Visual Question Answering   Few-Shot Learning   Multimodal Models   Flamingo   Semantic Alignment

👥 Authors

Chethmi Nayanathara   Corresponding

🗺 Sri Lanka
🏛 University of Moratuwa
✉ chethmi.21@cse.mrt.ac.lk

Uthayasanker Thayasivam

🗺 Sri Lanka
🏛 University of Moratuwa
✉ rtuthaya@cse.mrt.ac.lk

ⓘ Additional Information

Submission Type

Full Paper (Presentation and Publication)

Submission Date

2025-10-20 21:35:51

Last Updated

2025-10-20 21:35:51

📄 **Files**

Download Paper
236.84 KB

Update Paper

Withdraw Paper

← Back to List