**Title: A Novel Testing Framework for Vision Models Using Bayesian Network**

**Abstract:**

**Preamble:** Vision models (VMs), particularly those deployed for classification tasks in supervised learning, have become game-changers in critical domains, including autonomous vehicle navigation, healthcare diagnostics, and security surveillance systems. These models excel at processing and interpreting complex visual data, which is essential for applications ranging from navigating and interpreting road conditions to enhancing diagnostic accuracy and monitoring secure areas.

requiring precise visual data interpretation, such as autonomous vehicle systems, medical image analysis, and security monitoring. These models are critical in both understanding and reacting to complex environments.

**Challenge:**  The deployment of VMs in real-world environments requires rigorous testing across various settings to ensure robustness, from local to global level. This involves a systematic breakdown and summarising errors to gain clear insights into areas where improvement is necessary.

**Local-level robustness:** Local-level robustness assesses the model accuracy within specific categories. For example, in autonomous driving, this might involve consistently recognizing stop signs, regardless of weather conditions or angles; in healthcare, it means identifying specific tumor types accurately across various imaging technologies. Error summarization at this level records and analyse errors that occur within these specific labelled scenarios, helping to refine the model’s accuracy within each category.

**Global-level Robustness:**  Global robustness, on the other hand, evaluates the model effectiveness across a broader range of conditions and tasks. For instance, in autonomous driving, it ensures accurate recognition of stop signs under diverse road conditions and surrondings; similarly, in healthcare, it verifies the model's ability to identify tumors across different imaging techniques and patient demographics. Error summarization in this context aims to analyse model  performance across different scenarios, with the goal of improving its ability to generalise effectively.

**Gap:** Existing methods lack comprehensive assessments of VMs across different real-world conditions and do not effectively identify specific vulnerabilities.

​​

**MyResearch:**  This research introduces a groundbreaking testing framework that employs a probabilistic approach through Bayesian networks to systematically assess vision model robustness at both local and global scales. This framework includes detailed error summarization for precise evaluation at every level. Bayesian network combines prior knowledge with new real-world observations, making our evaluations more trustworthy. This research effectively handles uncertainty by systematically updating these evaluations, ensuring our framework remains robust and reliable in providing accurate estimates.

**Abstract:**

**Introduction:**

Vision models (VMs), particularly in supervised classification tasks, are critical in high-stakes domains such as autonomous driving, medical diagnostics, and security systems. These models play a pivotal role in interpreting complex visual data, thus enhancing capabilities in analysing road conditions, diagnosing diseases, and ensuring area security.

**Problem Statement**: The real-world deployment of VMs requires exhaustive testing to assess their robustness. This involves an in-depth error analysis and systematic error summarization at local and global scales to identify critical areas for enhancement.

**Local-level Robustness:** At the local level, robustness testing focuses on model accuracy within specific categories. For example, this could involve to ensure the consistent recognition of stop sign under various weather conditions in autonomous driving or identify specific tumour using different imaging techniques in healthcare. Error analyses and summarization at this stage are aimed to refine accuracy within these targeted scenarios.

**Global-level Robustness:** At the global levec, robustness involves assessing the model’s performance across a broad range of environmental conditions and tasks. This includes to validate the consistent recognition of stop sign across diverse road conditions and ensure accurate tumour detection across varied patient demographics and imaging techniques. The goal at this level is to improve the model’s ability to generalize effectively across different scenarios.

**Research Gap**: Existing frameworks do not adequately assess VMs across diverse conditions or effectively pinpoint specific areas of vulnerability.

**My Research:** This research proposes a novel framework to systematically test the robustness of VMs at both the local and global levels, incorporating comprehensive error reporting to detail each assessment. The integration of Bayesian methods enhances this framework by utilizing both prior knowledge and new empirical data to refine evaluations continuously. This approach not only improves the reliability of the assessments but also ensures the model adapts dynamically to new data, maintaining the accuracy and dependability of the system.

**Preamble:**

Vision models (VMs), particularly in supervised classification tasks, are critical in high-stakes domains such as autonomous driving, medical diagnostics, and security systems.

**Challenge**: The real-world deployment of VMs require exhaustive testing to assess their robustness. This involves detailed error analysis at local and global level to identify critical areas where improvement is necessary.

**Local and Global Level Robustness:** At the local level, testing focuses on model robustness within defined categories/classes. In contrast, global level assesses the generalization ability of model on broader range of scenarios.

**Research Gap**: Existing VMs testing frameworks lack  to merge prior knowledge  dynamically calculate the probability  lack comprehensive evaluation of VMs across diverse conditions or effectively pinpoint specific areas of vulnerability.

**My Research:**

This research introduces a comprehensive framework designed to enhance the testing and evaluation of VMs through a structured, four-stage process. The framework begins with the **Specification** module, where necessary specifications are clearly defined to guide the entire testing process. Following specifications, the **Sampling & Test Case Generation** module selects relevant test cases from a broader input test set, ensuring that only valid and applicable cases are forwarded for testing. The core of the framework, **Testing & Probabilistic Graph** module, then conducts robustness assessments at both local and global levels. This module utilizes probabilistic approach of baysian network to produce a graphical representation of model robustness, providing a visual and quantitative analysis of performance across different scenarios. Finally, the **Error Summarization** stage compiles these findings into an actionable format, offering graphical error reports and recommendations for model refinement. This framework systematically assesses VMs' robustness and integrates continuous feedback loops for ongoing improvement and accuracy in evaluations.

**Title: A Novel Testing Framework for Vision Models Using Bayesian Network**

**Abstract:**

**Preamble:**

Vision models (VMs), particularly in supervised classification tasks, are critical in high-stakes domains such as autonomous driving, medical diagnostics, and security systems.

**Challenge**: Real-world deployment of VMs requires rigorous testing for robustness, especially given the variety of operational environments. This requires precise error analysis at both local and global levels to pinpoint specific areas for improvement.

**Local and Global Level Robustness:** At the local level, testing focuses on model robustness within defined categories/classes. In contrast, global level assesses the generalization ability of model on broader range of scenarios.

**Research Gap**:

Existing VMs testing methods lack cohesive, structured framework that integrates detailed probabilistic assessments with thorough error summarization and visualization. Current practices also fail to present a continuous graphical representation of robustness at both local and global levels, essential for precise and actionable model improvements. Our proposed framework fills this gap by introducing a novel, four-stage process that systematically enhances the evaluation and visualization of VMs robustness at every stage.

**My Research:**

This research introduces a comprehensive testing framework designed to enhance  the evaluation of VMs through a structured, four-stage process. The framework begins with the **Specification** module, where necessary specifications are clearly defined to guide the entire testing process. Following specifications, the **Sampling & Test Case Generation** module selects relevant test cases from a broader input test set, ensuring that only valid and applicable cases are forwarded for testing. The core of the framework, **Testing & Probabilistic Graph** module, then conducts robustness assessments at both local and global levels. This module utilizes probabilistic approach of baysian network to produce a graphical representation of model robustness, providing a visual and quantitative analysis of performance across different scenarios. Finally, the **Error Summarization** stage compiles these findings into an actionable format, offering graphical error reports and recommendations for model refinement. This framework systematically assesses VMs' robustness and integrates continuous feedback loops for ongoing improvement and accuracy in evaluations.

**Abstract:**

**Preamble:**

Vision models (VMs), particularly in supervised classification tasks, are critical in high-stake domains such as autonomous driving, medical diagnostics, and security systems.

**Challenge**: Real-world deployment of VMs requires rigorous testing to check robustness, especially given the variety of operational environments. This requires precise error analysis at both local and global levels to pinpoint specific areas for improvement.

**Local and Global Level Robustness:** At the local level, testing focuses on model robustness within defined categories/classes. In contrast, global level assesses the generalization ability of model on broader range of scenarios.

**Research Gap**:

Current testing methods for vision models, such as Neuron Coverage (NC), k-Multisection Neuron Coverage (KMNC), Neuron Boundary Coverage (NBC), Strong Neuron Activation Coverage (SNAC), Modified Condition/Decision Coverage (MC/DC), Likelihood-based Surprise Adequacy (LSA), and Distance-based Surprise Adequacy (DSA), primarily focus on neural network specifics. While these metrics are valuable for measuring certain aspects of model behavior, they often do not comprehensively assess robustness in varied real-world conditions. This narrow focus overlooks the practical operational robustness needed to ensure reliability across diverse environments, leaving a significant gap in their thorough evaluation.

**My Research:**

This research introduces a comprehensive testing framework designed to enhance  the evaluation of VMs through a structured, four-stage process. The framework begins with the **Specification** module, where necessary specifications are clearly defined to guide the entire testing process. Following specifications, the **Sampling & Test Case Generation** module selects relevant test cases from a broader input test set, ensuring that only valid and applicable cases are forwarded for testing. The core of the framework, **Testing & Probabilistic Graph** module, then conducts robustness assessments at both local and global levels. This module utilizes probabilistic approach of baysian network to produce a graphical representation of model robustness, providing a visual and quantitative analysis of performance across different scenarios. Finally, the **Error Summarization** stage compiles these findings into an actionable format, offering graphical error reports and recommendations for model refinement. This framework systematically assesses VMs' robustness and integrates continuous feedback loops for ongoing improvement and accuracy in evaluations.

**Abstract:**

**Preamble:**

Vision models (VMs), particularly in supervised classification tasks, are critical in high-stake domains such as autonomous driving, medical diagnostics, and security systems.

**Challenge**: Real-world deployment of VMs requires rigorous robustness testing due to the diversity of operational environments. This challenge involves detailed error analysis at both local and global levels. At the local level, testing aims to asses model robustness within specific categories or classes, identifying precise areas for improvement. Conversely, at the global level, testing evaluates the  generalization ability of model across a wide range of scenarios. These dual layers of assessment are essential to ensure the model robustness under varied real-world conditions.

**Research Gap**:

Current testing methods for vision models, such as Neuron Coverage (NC), k-Multisection Neuron Coverage (KMNC), Neuron Boundary Coverage (NBC), Strong Neuron Activation Coverage (SNAC), Modified Condition/Decision Coverage (MC/DC), Likelihood-based Surprise Adequacy (LSA), and Distance-based Surprise Adequacy (DSA), primarily focus on neural network specifics. While these metrics are valuable for measuring certain aspects of model behavior, they often do not comprehensively assess robustness in varied real-world conditions. This narrow focus overlooks the practical operational robustness needed to ensure reliability across diverse environments, leaving a significant gap in their thorough evaluation.

**My Research:**

This research introduces a comprehensive testing framework designed to enhance  the evaluation of VMs through a structured, four-stage process. The framework begins with the **Specification** module, where necessary specifications are clearly defined to guide the entire testing process. Following specifications, the **Sampling & Test Case Generation** module selects relevant test cases from a broader input test set, ensuring that only valid and applicable cases are forwarded for testing. The core of the framework, **Testing & Probabilistic Graph** module, then conducts robustness assessments at both local and global levels. This module utilizes probabilistic approach of baysian network to produce a graphical representation of model robustness, providing a visual and quantitative analysis of performance across different scenarios. Finally, the **Error Summarization** stage compiles these findings into an actionable format, offering graphical error reports and recommendations for model refinement. This framework systematically assesses VMs' robustness and integrates continuous feedback loops for ongoing improvement and accuracy in evaluations.

**Abstract:**

**Preamble:**

Vision models (VMs), particularly in supervised classification tasks, are critical in high-stake domains such as autonomous driving, medical diagnostics, and security systems.

**Research Gap**: Current testing approaches for  VMs, such as Neuron Coverage (NC), k-Multisection Neuron Coverage (KMNC), Neuron Boundary Coverage (NBC), Strong Neuron Activation Coverage (SNAC), Modified Condition/Decision Coverage (MC/DC), Likelihood-based Surprise Adequacy (LSA), and Distance-based Surprise Adequacy (DSA), primarily focus on neuron coverage, which can be misleadingly satisfied with minimal data samples. This focus insufficiently captures the models’ robustness across varied real-world scenarios, leaving a crucial gap in comprehensive and practical robustness evaluation.

**Challenge**: Real-world deployment of VMs requires rigorous robustness testing due to the diversity of operational environments. This challenge involves detailed error analysis at both local and global levels. Local level define to asses the model robustness within specific categories or classes, to identify precise areas for improvement. Conversely, global level refers to the  generalization ability of model across a wide range of scenarios. These dual layers of assessment are essential to ensure the model robustness under varied real-world conditions.

**My Research:** This research introduces a comprehensive testing framework designed to enhance  the evaluation of VMs through a structured, four-stage process. The framework begins with the **Specification** module, where necessary specifications are clearly defined to guide the entire testing process. Following specifications, the **Sampling & Test Case Generation** module selects relevant test cases from a broader input test set, ensuring that only valid and applicable cases are forwarded for testing. The core of the framework, **Testing & Probabilistic Graph** module, then conducts robustness assessments at both local and global levels. This module utilizes probabilistic approach of baysian network to produce a graphical representation of model robustness, providing a visual and quantitative analysis of performance across different scenarios. Finally, the **Error Summarization** stage compiles these findings into an actionable format, offering graphical error reports and recommendations for model refinement. This framework systematically assesses VMs' robustness and integrates continuous feedback loops for ongoing improvement and accuracy in evaluations.

**Title: A Novel Testing Framework for Vision Models Using Bayesian Network**

**Abstract:**

**Preamble:**

Deep learning models, particularly vision models (VMs), are critical in high-stake domains such as autonomous driving, medical diagnostics, and security systems. The real-world deployment of VMs requires rigorous robustness testing because of diverse environmental conditions.

**Challenge**: The current testing approaches for VMs primarily focus on neuron coverage. While this is a critical metric, it alone does not ensure comprehensive coverage of all corner cases, leaving a gap in the overall evaluation of the model’s robustness.

**Proposed Framework:** My research develops an extensive testing framework designed to enhance  the evaluation of VMs through a structured, **five-stage process.**

* **Specification Module:** This initial stage focuses on clearly defining all necessary properties of the system. This detailed specification guides the entire testing process and ensures all aspects of the system are covered.
* **Sampling**: The second stage involves gathering all relevant samples that are necessary to test the system thoroughly. This includes a comprehensive collection of scenarios the system may encounter.
* **Test Case Generation:** In the third stage, properties defined in the Specification module are applied to the collected samples. Test cases are then generated based on these specifications. For example, in testing an autonomous car, factors like dust, noise, rain, and night conditions are considered to evaluate the system’s performance under these conditions.
* **Testing & Probabilistic Graph:** The fourth stage begins with testing the generated test cases to validate their effectiveness. After testing, robustness assessments are conducted both locally and globally. Locally, the robustness of the model is evaluated within individual categories or classes to pinpoint weaknesses. Globally, the model’s performance is assessed across various categories to enhance its generalization capabilities across different scenarios. Errors are systematically recorded for later analysis. A probabilistic approach using Bayesian networks, combined with solid mathematical formulations, is integrated to provide a comprehensive visual and quantitative analysis of the model’s performance at both local and global levels.
* **Error Summarization:** The final stage compiles and analyses the recorded errors, producing actionable graphical error reports and recommendations for VMs refinement.