PhD in INGEGNERIA DELL'INFORMAZIONE / INFORMATION TECHNOLOGY - 41st cycle

Research Area n. 1 - Computer Science and Engineering

THEMATIC Research Field: MULTIMODAL ANOMALY DETECTION FOR INDUSTRIAL PRODUCTS QUALITY CONTROL

Monthly net income of PhDscholarship (max 36 months)
1500.0 In case of a change of the welfare rates during the three-year period, the amount could be modified.

Context of the research activity	
Motivation and objectives of the research in this field	Ensuring consistent product quality is one of the most crucial challenges in modern manufacturing. Automated visual inspection systems are increasingly deployed to detect defects in industrial products, reducing human error and improving throughput and consistency. However, most current approaches to anomaly detection (AD) in quality control rely solely on visual data. These systems typically learn the appearance of defect-free products and identify anomalies as deviations from this learned representation. While highly effective under constrained conditions, visual-only approaches fail to leverage the abundance of contextual and semantic information often available in industrial processes—such as technical documentation, product specifications, operator notes, or textual descriptions of acceptable variations. This proposal introduces a new research direction—Multimodal Anomaly Detection (MAD)—that integrates both visual and textual information to perform more robust and explainable quality control. The central hypothesis is that models capable of jointly understanding visual and textual modalities will achieve superior anomaly detection and localization performance in specific industrial settings, where classic unsupervised image-only approaches fall short. Despite major advances in multimodal learning and

	vision–language models (VLMs), their potential for industrial anomaly detection remains largely unexplored. Moreover, there is a critical lack of benchmark datasets and evaluation frameworks that capture the multimodal aspects of anomaly detection tasks. Current academic benchmarks (e.g., MVTec AD, VisA, BTAD, Real IAD, MVTec AD 2, MVTec LOCO) only evaluate models on visual data, preventing meaningful assessment of how models reason about textual cues. The proposed PhD, proposed in collaboration with Pirelli Tyre S.P.A. aims to bridge this gap by defining the new task of multimodal anomaly detection, developing benchmark datasets for it, evaluating existing models, and creating new architectures that integrate multimodal reasoning into industrial quality control. Beyond text and images, the project will also explore extensions to other sensory modalities, paving the way for the next generation of AI-based industrial inspection systems.
Methods and techniques that will be developed and used to carry out the research	The project aims to advance multimodal anomaly detection for industrial products quality control process. First, it builds a framework to create novel benchmarks by converting existing visual AD dataset. The project should develop a structured pipeline: normal images receive textual descriptions that encode expected texture, color, geometry, and materials; targeted inspection prompts require reasoning about object composition, part-level conditions, defect types, and likely severity; annotations are produced with automated captioning, rule-based template filling, and expert curation. The final datasets will always include pixel-level ground truth masks for defect localization. This benchmark creation phase will be complemented by an application-oriented tyre dataset (tread/sidewall/bead, varied lighting and production settings) created with the company support to mirror real inspection scenarios. Once the evaluation scenarios will be ready the student is required to systematically test contemporary multimodal approaches—zero-/few-shot. anomaly detection methods, Large Vision Language/Any-to-Any models, and Vision-Question-Answering style systems—emphasizing joint text–image reasoning. A full setup of evaluation metrics

	both from anomaly detection/localization perspective and from the quality of generated explanations and responses will be adopted. Insights from this investigation will help the development of a new model that fuses visual and textual embeddings to emit a global anomaly score and a dense anomaly map; training explores a normal-only regime (text-guided feature alignment to enhance generalization) and a fully supervised regime with domain adaptation for unseen data, validated on the tyre datasets to test robustness and transfer to industry. Finally, the framework broadens to additional sensing beyond RGB (including depth and related non-RGB signals), where alternative fusion strategies are compared to quantify robustness gains under real-world challenges such as reflective materials, low illumination, and occlusions.
Educational objectives	The primary educational objective for this Ph.D. research is to train the student in the comprehensive process of scientific research and technological development within the field of Artificial Intelligence, specifically applied to industrial quality control. This includes gaining a deep understanding of how to conduct original, impactful research by defining a novel problem—Multimodal Anomaly Detection (MAD)—and systematically addressing its core challenges. A critical goal is to publish research findings in relevant, top-tier AI conferences and journals, thereby contributing new knowledge to the multimodal learning and vision anomaly detection communities. The student will be expected to develop a strong theoretical foundation in vision-language models and anomaly detection, coupled with the practical skill of designing and implementing innovative AI solutions, such as creating a new multimodal AD architecture and corresponding training procedures. Furthermore, a significant practical objective is to develop an application-oriented proof of concept for a multimodal anomaly detection method, demonstrating its effectiveness in an automated quality control process using real-world industrial data, specifically with the partner company's tyre datasets. Finally, the student is expected to demonstrate the general competencies of a Ph.D. graduate, including

	the ability to independently manage a large-scale research project, build and evaluate benchmark datasets, critically assess the strengths and limitations of existing models, and bridge the gap between academic research and industrial deployment.
Job opportunities	This Ph.D. project is highly relevant to current and future industry demands, positioning the student for excellent job opportunities upon graduation. The interdisciplinary focus on advanced AI, multimodal learning, and industrial quality control makes the graduate a prime candidate for roles such as Research Scientist (AI related fields), Research Engineer, or Machine Learning Engineer in sectors utilizing automated inspection. Specific sectors include manufacturing, automotive (especially with the direct collaboration with Pirelli Tyre S.p.A.), electronics, and aerospace. The expertise gained in developing models robust and exploitable in real-world industrial environments is highly sought after by major technology companies and specialized industrial AI startups looking to deploy the next generation of inspection systems.
Composition of the research group	1 Full Professors 0 Associated Professors 2 Assistant Professors 1 PhD Students
Name of the research directors	Prof. Matteo Matteucci, Dr. Stefano Samele

Contacts
stefano.samele@polimi.it, matteo.matteucci@polimi.it

Additional support - Financial aid per PhD student per year (gross amount)	
Housing - Foreign Students	--
Housing - Out-of-town residents	--

Scholarship Increase for a period abroad	
Amount monthly	750.0 €
By number of months	6

Additional information: educational activity, teaching assistantship, computer availability, desk availability, any other information

EDUCATIONAL ACTIVITIES (purchase of study books and material, including computers, funding for participation in courses, summer schools, workshops and conferences): financial aid per PhD student.

TEACHING ASSISTANTSHIP: availability of funding in recognition of supporting teaching activities by the PhD student. There are various forms of financial aid for activities of support to the teaching practice. The PhD student is encouraged to take part in these activities, within the limits allowed by the regulations.

COMPUTATIONAL RESOURCES: the student will be granted access to the AIRLab servers and resources in order to support the research activity. In case of need, Pirelli Tyre S.p.A. will supply further access to computing resources.