

OnSafety Vision AI (OnVI): Um Sistema de Detecção de EPIs e não conformidades via Ensemble de Redes Neurais e Super Resolução em Arquitetura Híbrida e Adaptativa

Luiz M. F. de Lima², Jonathan C. Kuspil^{1,2}, Karina Y. Kimura^{1,2}, Caio Flexa²,
Renan Ceratto², Gislaine Camila L. Leal^{1,2}, Edwin V. C. Galdamez^{1,2}

¹Universidade Estadual de Maringá – Maringá – PR – Brasil

²OnSafety – Maringá, PR – Brasil

{luiz.m.franca1, jonathankuspil, kimurakarina23}@gmail.com

caio.rodrigues@icen.ufpa.br, ceratto@onsafety.com.br

{gclleal, evcgaldamez}@gmail.com

Abstract. *Manual PPE monitoring is inefficient given the complexity of industrial environments. This paper proposes OnVI (OnSafety Vision AI), a Computer Vision solution based on an Adaptive Hybrid Architecture, integrating privacy and scalability. The system orchestrates an Ensemble of YOLOv4 PPE inference networks (Generalist and Specialist), pose estimation, and Super Resolution to overcome occlusion and distance limitations. This strategy aims to compensate for the limitations of isolated detectors via the robustness of the set, maximizing precision in identifying compliances and non-compliances and minimizing false positives, without neglecting the infrastructure flexibility required by modern industry.*

Resumo. *O monitoramento manual de EPIs apresenta-se ineficiente frente à complexidade do ambiente industrial. Este trabalho propõe o OnVI (OnSafety Vision AI), uma solução de Visão Computacional baseada em uma Arquitetura Híbrida e Adaptativa que integra privacidade e escalabilidade. O sistema orquestra um Ensemble de redes YOLOv4 de inferência de EPIs (Generalistas e Especialistas), estimativa de pose e Super Resolução para superar limitações de oclusão e distância. Essa estratégia visa compensar as limitações de detectores isolados por meio da robustez do conjunto, maximizando a precisão na identificação de conformidades e não conformidades e minimizando falsos positivos, sem negligenciar a flexibilidade de infraestrutura exigida pela indústria moderna.*

1. Contexto

A segurança industrial enfrenta desafios críticos quanto à adesão aos Equipamentos de Proteção Individual (EPIs), refletidos nos 732.751 acidentes em 2023 e custos bilionários registrados desde 2012 apenas no Brasil [Smartlab 2026]. Apesar da obrigatoriedade imposta pela NR-06 [Brasil 2022], fatores como desconforto e pressão por produtividade levam à negligência [Wong et al. 2020], agravando-se pela ineficiência estrutural da fiscalização visual humana. Ademais, esse método tradicional é suscetível ao “Efeito

Hawthorne” [McCambridge et al. 2014], fenômeno em que a conformidade ocorre apenas durante a presença observável do fiscal, gerando pontos cegos operacionais propícios a acidentes assim que a supervisão se ausenta.

Nesse cenário, a Visão Computacional surge como ferramenta para fiscalização contínua, porém o estado da arte atual carece de robustez para o ambiente fabril. Suas soluções frequentemente baseiam-se em detectores únicos ou arquiteturas simplificadas [Delhi et al. 2020, Protik et al. 2021, Gallo et al. 2022, Do et al. 2023], falhando em cobrir requisitos complexos como a garantia de privacidade das imagens, a variação na qualidade de captura e a validação espacial da ergonomia do EPI. Essa lacuna tecnológica demanda sistemas capazes de lidar com a complexidade real da indústria.

Para superar tais limitações, este trabalho propõe o **OnVI (OnSafety Vision AI)**¹, consistindo em uma Arquitetura Híbrida e Adaptativa. O sistema inova ao implementar um serviço de detecção composta (*Ensemble Learning*) [Dietterich 2000, Ganaie et al. 2022], orquestrando redes neurais generalista, especialistas e de pose, integradas à Super Resolução. Essa estratégia visa compensar as fraquezas de detectores isolados, maximizando a precisão na identificação de conformidades e minimizando falsos positivos, sem negligenciar a flexibilidade de infraestrutura exigida pela indústria moderna, consolidando o OnVI como um Sistema de Informação (SI) de suporte à decisão.

2. Processo Adotado

As etapas adotadas para o desenvolvimento da solução OnVI seguiram uma abordagem experimental iterativa, inspirada em práticas ágeis de desenvolvimento de software e adaptada para o ciclo de vida do projeto. O processo foi estruturado em seis fases distintas, detalhadas a seguir:

1. **Entendimento do Problema e Levantamento de Requisitos:** Nesta fase inicial, foram mapeadas as complexidades do ambiente industrial e definidos os requisitos funcionais e não funcionais, considerando a obrigatoriedade da privacidade das imagens internas da empresa e a possível necessidade de processamento em nuvem. Identificaram-se também as classes (conformidades) e contraclasses (não conformidades de EPI) de interesse para as redes, abrangendo as categorias de EPIs, essenciais para a regra de negócio de conformidade.
2. **Curadoria de Dados e Prototipação:** Envolveu a criação de um *dataset* heterogêneo contendo cerca de 33.661 imagens (mais de 94.000 amostras de classes) de ambientes reais e simulados. Simultaneamente, definiu-se o fluxo de processamento e a criação de protótipos para validar a viabilidade da arquitetura de detecção em diversos cenários.
3. **Treinamento e Modelagem da IA:** A etapa consistiu na estratégia de consolidar módulos de detecção distintos, conforme ilustrado na Figura 1, operando em conjunto ou paralelo para cobrir e corrigir as detecções das classes e contraclasses de interesse. Englobando também redes especializadas para inferir bounding boxes – caixas delimitadoras de objetos – de pessoas e EPIs e estimar poses.
4. **Implementação:** Dedicada ao desenvolvimento do OnVI aplicando princípios de engenharia de software moderna. O DDD [Evans 2003] delimitou os contextos do domínio de segurança, enquanto o SOLID [Martin 2003] garantiu o

¹OnSafety: Empresa de software para SST. Informações sobre o uso de IA disponíveis em: <https://onsafety.com.br/a-inteligencia-artificial-ia-na-sst/>. Acesso em: 5 fev. 2026.

baixo acoplamento necessário para orquestrar o *Ensemble* de redes de forma modular. Arquiteturalmente, o sistema foi concebido como um ecossistema de Arquitetura Híbrida e Adaptativa, utilizando microsserviços de coleta de imagens, comunicação assíncrona e detecção de EPIs.

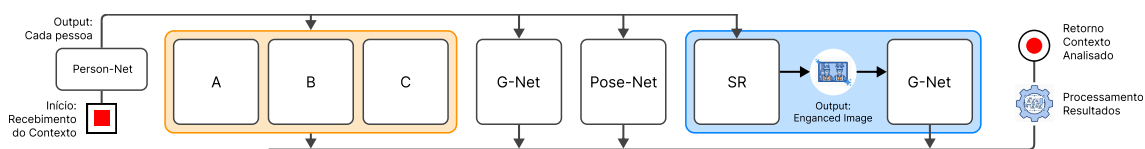


Figura 1. Fluxo de inferência: recorte de pessoas seguido pelo pipeline de redes, processamento dos resultados e retorno do contexto analisado.

5. **Teste, Validação e Evolução:** Esta etapa envolve a realização de testes *End-to-End* aferindo métricas estatísticas (*F1-Score* e *Recall*) e de latência para validar a viabilidade em ambientes simulados e reais, configurando parâmetros para cada perfil de cliente visando melhorias contínuas.
6. **Deploy e Feedback:** Implantação do sistema nos ambientes industrial e de nuvem, seguida da coleta de *feedback* acerca das experiências dos usuários finais.

3. Solução

A solução desenvolvida, *OnVI (OnSafety Vision AI)*, fundamenta-se em um serviço de inteligência artificial que orquestra módulos neurais especializados. O fluxo de inferência, ilustrado pela Figura 1, inicia-se pela iteração do **Contexto** — uma lista de imagens que representa um intervalo temporal do ambiente — onde a **Person-Net** isola os trabalhadores na cena, gerando recortes individuais que alimentam o pipeline subsequente. A análise de conformidade é realizada paralelamente por módulos que utilizam as redes **G-Net** (Generalista), **ABC-Net** (Especialista) e **Pose-Net**, baseadas na arquitetura *YOLOv4 (You Only Look Once, version 4)* [Bochkovskiy et al. 2020]. Embora as redes **G-Net** e **ABC-Net** sejam capazes de inferir classes e contraclasses de EPIs, a **ABC-Net** distingue-se por adotar uma abordagem de *Ensemble Learning*, sendo composta por redes treinadas a partir de subconjuntos disjuntos contidos no conjunto global de classes, potencializando seu desempenho em comparação à **G-Net**. Por fim, a **Pose-Net** estima a geometria corporal para validar espacialmente as detecções (e.g., confirmando o posicionamento anatômico correto dos EPIs), enquanto o módulo condicional de Super Resolução (**SR**) via *Real-ESRGAN* [Wang et al. 2021] atua na reconstrução de imagens degradadas pela distância, submetendo-as a uma nova inferência pela rede **G-Net**.

A eficácia dessa arquitetura foi validada por meio de experimentos comparativos. Em cenários padrão, os dados demonstram que a estratégia de especialização da **ABC-Net** superou a abordagem generalista, elevando o *F1-Score* médio de 0.67 (**G-Net**) para 0.72 (**ABC-Net**) e aumentando o *Recall* de 0.57 para 0.64. Isso indica uma redução significativa nos falsos negativos e falsos positivos, garantindo maior sensibilidade à ausência de equipamentos. O impacto da inovação torna-se ainda mais evidente em cenários de longa distância (10 a 40 metros), onde a resolução nativa é insuficiente. A introdução da Super Resolução permitiu que a sua aplicação seguida da rede **G-Net** sustentasse um *F1-Score* de 0.6 e um *Recall* de 0.46, desempenhos consideravelmente superiores à **G-Net** padrão, 0.54 e 0.38 respectivamente, em distâncias maiores. Esses resultados confirmam que a essa composição de redes é indispensável para a aplicabilidade na industrial real.

Para viabilizar a operação desses modelos em escala fabril, o sistema foi encapsulado em microsserviços orientados a eventos, conforme apresentado na Figura 2. A comunicação é orquestrada por uma mensageria (RabbitMQ), por meio de uma fila de mensagens, que promove o desacoplamento estrito entre a coleta e a inferência. No ambiente local (Figura 2a), microsserviços dedicados garantem a aquisição de imagens e a orquestração local com uma gestão resiliente dos Contextos, assegurando que os dados sejam preservados e disponibilizados assincronamente. Essa estratégia é o alicerce que permite ao microsserviço de processamento operar de forma agnóstica à infraestrutura, consumindo as requisições tanto em uma implantação local quanto em nuvem sem necessidade de refatoração.

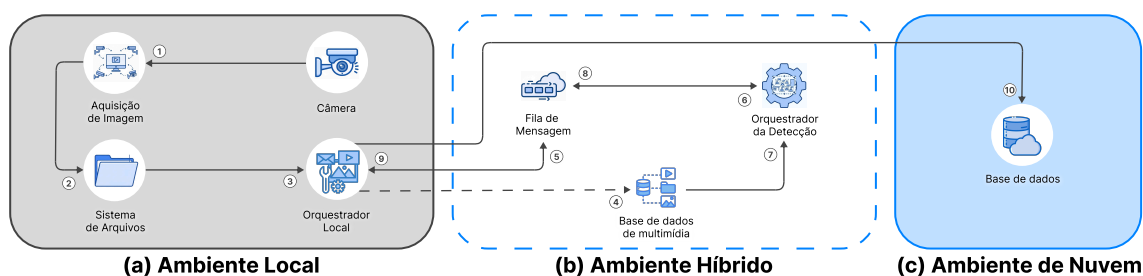


Figura 2. Arquitetura do OnVI, ilustrando a orquestração dos microsserviços e o fluxo de dados entre os ambientes Local, Híbrido e de Nuvem.

Diferenciando-se das soluções do estado da arte ao adotar uma Arquitetura Híbrida e Adaptativa, conforme Figura 2b, o OnVI oferece flexibilidade de infraestrutura para atender requisitos rigorosos de privacidade e custo. A arquitetura suporta dois modos de operação: (1) **On-Premise Deployment**: onde todo o processamento, incluindo a inferência das redes neurais pelo Orquestrador de Detecção, ocorre localmente no servidor do cliente. Neste cenário, a persistência de mídia é unificada no próprio Sistema de Arquivos do servidor local, garantindo que imagens sensíveis jamais trafeguem para redes externas, assegurando total conformidade com a LGPD [Brasil 2018]; e (2) **Cloud-Offloaded Deployment**: cujo core local realiza apenas a coleta e o envio de Contextos para processamento em nuvem, reduzindo o investimento em hardware local. Independentemente do modo de operação, os metadados analíticos resultantes são centralizados na Base de Dados em nuvem (Figura 2c), tornando as informações acessíveis para consumo via API pela plataforma de gestão.

Essa combinação de inteligência artificial robusta e arquitetura de software flexível ataca diretamente os problemas de gestão de segurança. Ao automatizar o monitoramento, a solução mitiga o “Efeito Hawthorne”, eliminando a dependência da presença física do fiscal, e ilumina os “pontos cegos” operacionais. O resultado final é entregue aos gestores (SESMT) por meio de *dashboards* analíticos e alertas proativos, transformando a cultura de segurança de reativa para preventiva.

Agradecimentos

Agradecemos a SETI - PR pelo financiamento do Projeto de Inovação Tecnológica aprovado no Edital CP 01/2023 - AGEUNI - SETI PR, Termo de Convênio N° 023/2024 - SETI/UGF/FADEC-UEM/UEM/Lean Tecnologia e Engenharia LTDA. ePROTOCOLO PR N° 23.379.733-2.

Referências

- Bochkovskiy, A., Wang, C.-Y., and Liao, H.-Y. M. (2020). Yolov4: Optimal speed and accuracy of object detection.
- Brasil (2018). Lei nº 13.709, de 14 de agosto de 2018. *Diário Oficial da República Federativa do Brasil*.
- Brasil (2022). NR 06 - EQUIPAMENTOS DE PROTEÇÃO INDIVIDUAL - EPI. Acesso em: 21 de janeiro de 2026.
- Delhi, V., Lal, S., and Thomas, A. (2020). Detection of personal protective equipment (ppe) compliance on construction site using computer vision based deep learning techniques. *Frontiers in Built Environment*, 6.
- Dietterich, T. G. (2000). Ensemble methods in machine learning. In *Multiple Classifier Systems*, pages 1–15, Berlin, Heidelberg. Springer Berlin Heidelberg.
- Do, M.-T., Kim, D. T., Ha, M.-H., Chen, O. T.-C., Chinh, N., and Tran, A. (2023). An effective method for detecting personal protective equipment at real construction sites using the improved yolov5s with siou loss function. pages 430–434.
- Evans, E. (2003). *Domain-Driven Design: Tackling Complexity in the Heart of Software*. Pearson Education.
- Gallo, G., Di Rienzo, F., Garzelli, F., Ducange, P., and Vallati, C. (2022). A smart system for personal protective equipment detection in industrial environments based on deep learning at the edge. *IEEE Access*, 10:110862–110878.
- Ganaie, M., Hu, M., Malik, A., Tanveer, M., and Suganthan, P. (2022). Ensemble deep learning: A review. *Engineering Applications of Artificial Intelligence*, 115:105151.
- Martin, R. (2003). *Agile Software Development: Principles, Patterns, and Practices*. Alan Apt series. Pearson Education.
- McCambridge, J., Witton, J., and Elbourne, D. R. (2014). Systematic review of the hawthorne effect: New concepts are needed to study research participation effects. *Journal of clinical epidemiology*, 67(3):267–277.
- Protik, A., Hossain, A., and Siddique, S. (2021). Real-time personal protective equipment (ppe) detection using yolov4 and tensorflow.
- Smartlab (2026). Observatório de segurança e saúde no trabalho. Acesso em: 21 janeiro de 2026.
- Wang, X., Xie, L., Dong, C., and Shan, Y. (2021). Real-esrgan: Training real-world blind super-resolution with pure synthetic data. In *International Conference on Computer Vision Workshops (ICCVW)*.
- Wong, T. K. M., Man, S. S., and Chan, A. H. S. (2020). Critical factors for the use or non-use of personal protective equipment amongst construction workers. *Safety Science*, 126:104663.