Inspeção de Falhas em Máscaras de Solda Utilizando Vision Transformers para Melhoria da Fabricação

[1]**Walter Jonas de S. Viana e** [1]**Felipe Gomes de Oliveira**

[1]Instituto de Ciências Exatas e Tecnologia –
Universidade Federal do Amazonas (ICET/UFAM) – Itacoatiara – Amazonas – Brasil

walter.viana@ufam.edu.br, felipeoliveira@ufam.edu.br

Resumo. *O controle de qualidade é vital na fabricação moderna. Este artigo aborda a inspeção de chips GSM durante a soldagem em substratos de PCB. Propomos um método baseado em Vision Transformers, que utiliza a autoatenção para aprender características visuais espaciais e hierárquicas. Nossa abordagem alcançou 95,83% de precisão em experimentos reais e simulados. Mesmo sob ruído e desfoque, a precisão permaneceu alta: 94,94% (Sal & Pimenta) e 94,17% (Gaussiano). Os resultados confirmam a robustez e o potencial industrial do método.*

1. Introdução

A crescente demanda por eletrônicos tornou a produção de circuitos integrados (IC) mais complexa devido ao alto volume [Wu et al. 2022]. Garantir qualidade é essencial para o sucesso [Silva et al. 2022b]. A inspeção visual tradicional feita por operadores humanos é suscetível à fadiga e distrações, comprometendo a confiabilidade [Silva et al. 2022a].

A Visão Computacional permite a Inspeção Visual Automática (AVI) com descritores como HOG e LBP e classificadores como SVM. Redes neurais como CNNs, Autoencoders e ResNet são aplicadas na detecção de falhas [Swathi et al. 2023]. Esses modelos aumentam a precisão e reduzem a intervenção humana.

Este artigo propõe uma abordagem com Vision Transformer para detectar falhas na máscara de solda em chips GSM. O modelo aprende padrões visuais nas conexões dos chips. Ao contrário de estudos sobre juntas de solda, focamos em falhas de máscara em regiões distintas. Resultados reais e simulados mostram alta acurácia e uso industrial.

Nossas principais contribuições são resumidas a seguir:

- Apresentamos uma abordagem baseada em Vision Transformer (ViT), técnica de ponta (SOTA) para detectar falhas de solda em chips. O método garante robustez ao analisar múltiplas máscaras por inspeção; A Figura 1 destaca o tamanho reduzido e a densidade das máscaras em chips GSM;
- Introduzimos também um conjunto de dados desafiador com mais de 1000 imagens de máscaras de solda extraídas de chips GSM, obtidas em ambiente industrial. Até onde sabemos, é o primeiro dataset desse tipo.

Além disso, as contribuições deste trabalho foram publicadas em uma conferência internacional relevante na área.

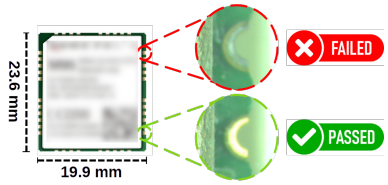


Figura 1. Ilustração de um chip GSM e sua máscara de solda.

2. Trabalhos Relacionados

Visão Computacional e Deep Learning tornaram a inspeção visual mais eficaz e confiável [Wang et al. 2023, Al Radi et al. 2023, Iglesias et al. 2021]. Pesquisas atuais visam detectar falhas em soldagem e máscaras de solda [Silva et al. 2022a, Heng 2022], integrando Visão Computacional e Aprendizado de Máquina com extração de características.

[Rocha et al. 2016] obteve 97,25% de acurácia com HOG e SVM na detecção de componentes SMC sob ruído. [Beuth et al. 2020] alcançou 88,0% usando atenção visual em wafers. [Swathi et al. 2023] aplicou YOLOv8 em PCBs (mAP_{50} de 0,955 e mAP_{50-95} de 0,516), evidenciando a eficácia da visão profunda.

Na inspeção de soldagem, [Silva et al. 2022a] utilizou CNN para detectar falhas em esferas de solda (98,009%). Já [Zhang et al. 2022] aplicou imagens de raio-X 3D em juntas de solda, com bom desempenho. A integração dessas técnicas reforça a confiabilidade em ambientes industriais.

O estudo mais relevante sobre máscaras de solda [Heng 2022] utilizou YOLOv5 para detectar delaminação, com 98,70% de precisão. A abordagem usou ROI para recorte e rotulagem, sendo eficaz na automação. Diferente de estudos baseados em CNNs, este trabalho aplica Vision Transformers testados sob ruído e desfoque. O método proposto alcançou alta acurácia e apresenta um novo dataset industrial de falhas em chips GSM.

3. Metodologia

Neste trabalho propomos uma abordagem baseada em Vision Transformers (ViT) para detectar falhas em máscaras de solda em chips GSM, rotulando imagens laterais ($l_1 \dots l_4$) conforme sua condição.

Problema [inspeção automatizada]: Seja $\boldsymbol{\mathcal{I}} = \{i_1, \dots, i_n\}$ um conjunto de imagens de regiões $\boldsymbol{\mathcal{L}}_j = \{l_1, \dots, l_4\}$ e rótulos $\boldsymbol{\mathcal{M}} = \{m_1, \dots, m_k\}$. Deseja-se classificar l_q em m_l.

3.1. Pré-Processamento

As imagens dos chips GSM ($\boldsymbol{\mathcal{I}}$) são cortadas para isolar regiões laterais de interesse (ROI), conforme:

$$\boldsymbol{\mathcal{L}} = \boldsymbol{\mathcal{I}}[y_0 : y_0 + h, x_0 : x_0 + w]. \tag{1}$$

Aqui, y_0 e $y_0 + h$ são os índices das linhas, e x_0 e $x_0 + w$ os das colunas. A Figura 2 mostra o recorte lateral com máscaras de solda na imagem.

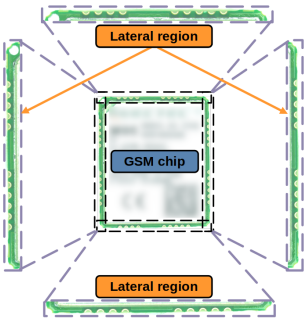


Figura 2. Segmentação de regiões laterais da imagem para aprendizado.

3.2. Arquitetura do Vision Tranformer

O Vision Transformer (ViT) é uma rede neural que adapta o transformer, criado para NLP, à classificação de imagens. ViTs igualam ou superam CNNs de ponta [Dosovitskiy et al. 2021]. A arquitetura usada, ilustrada na Figura 3, possui três etapas: divisão em blocos de imagem (*image patching*), incorporação de características (*embedding*) e codificador (*encoder*). Os detalhes são apresentados em seguida.

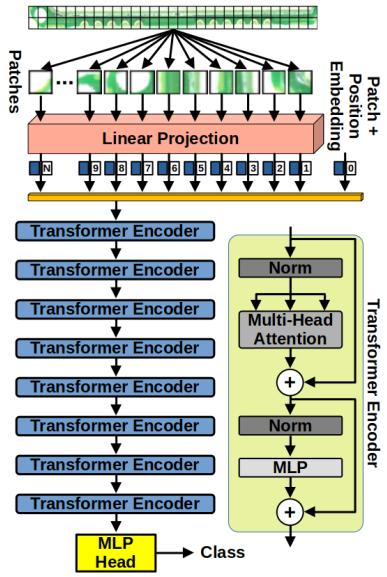


Figura 3. Arquitetura ViT proposta.

3.2.1. Divisão da imagem em patches

Uma imagem (l_j) de tamanho ($H \times W \times C$) é dividida em N patches não sobrepostos de $P \times P$, onde:

$$N = \frac{H.W}{P^2}, \quad X_i \in \mathbb{R}^{P^2.C}, \quad i = 1, 2, \ldots, N. \tag{2}$$

3.2.2. Incorporação dos patches

Os patches achatados são projetados linearmente em um espaço de embedding menor por uma matriz aprendível E:

$$Z_0^i = EX_i + e_{pos}^i, \tag{3}$$

Com $E \in \mathbb{R}^{D \times (P^2.C)}$ e $e_{pos}^i \in \mathbb{R}^D$, obtém-se a sequência de entrada do transformador.

$$Z_0 = [Z_0^1; Z_0^2; ...; Z_0^N] \in \mathbb{R}^{N \times D}. \tag{4}$$

3.2.3. Codificador

O codificador Transformer possui L camadas de autoatenção multi-cabeça (MSA), perceptron multi-camadas (MLP), normalização (LN) e conexões residuais.

MSA: Para cada camada l, computa-se:

$$A = \text{softmax}\left(\frac{QK^T}{\sqrt{D}}\right)V, \quad \text{com } Q = Z^{l-1}W^Q,\ K = Z^{l-1}W^K,\ V = Z^{l-1}W^V \tag{5}$$

As saídas das cabeças são concatenadas:

$$Z_{MSA} = \text{Concat}(\text{head}_1, ..., \text{head}_h)W^O \tag{6}$$

LN e MLP: Aplicam-se conexões residuais e normalização:

$$Z_l' = \text{LN}(Z^{l-1} + Z_{MSA}), \quad Z_l = \text{LN}(Z_l' + \text{MLP}(Z_l')) \tag{7}$$

com:

$$\text{MLP}(Z) = \text{GELU}(ZW_1 + b_1)W_2 + b_2 \tag{8}$$

Classificação: O token de classe z_L^0 gera a predição final:

$$\text{Output} = \text{softmax}(W_{cls}z_L^0) \tag{9}$$

onde W_{cls} é uma matriz de projeção aprendida.

A arquitetura ViT proposta extrai patches 16x16, achata e projeta para vetores de dimensão 8, e 4 cabeças de atenção. Tem 8 blocos codificadores transformer e um MLP com 512 e 256 unidades. O treinamento usa o otimizador Adam ($lr = 0{,}001$, batch 16, 200 épocas, eficaz para classificação sob diversas condições.

4. Experimentos

4.1. Configurações experimentais

Nesta seção, apresentamos os experimentos realizados para validar a proposta.

4.1.1. Detalhes de implementação

Nosso método foi implementado em TensorFlow, em estação Dell com CPU Xeon® Silver 4114 (2.20GHz), 128GB RAM e GPU RTX A4000 (16GB). As imagens foram capturadas com câmera Basler AcA5472-17uc e lente TS1614 F1.4 f50mm; patch size e projeção definem a codificação. O ViT foi treinado com $lr = 0{,}001$, batch 16, 200 épocas, ajustes via grid search e avaliação por validação cruzada 5-fold.

4.1.2. Dataset

Propomos o Solder Mask Inspection (SMI), com 1079 imagens extraídas de chips GSM. Aplicamos data augmentation (ajuste de brilho, flip, crop, cutout). Tamanho reduzido de 2770x210x3 para 1385x105x3. A Figura 4 mostra exemplos.

(a) **(b)**

Figura 4. Visualização de máscaras de solda defeituosas (4a) e corretas (4b).

4.2. Avaliação Experimental

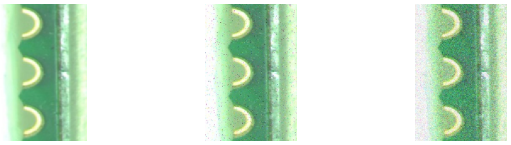
Os experimentos realizados avaliam a precisão da abordagem proposta para inspeção de máscaras de solda, comparando seis métodos: Vision Transformer (Nosso), CNN, CNN com Autoencoder (AE), CNN com Random Forest (RF), CNN com SVM e Autoencoder. Técnicas de deep learning foram escolhidas por sua eficácia na inspeção visual [Silva et al. 2022a][Hida et al. 2021][Dong et al. 2018][Kumaresan et al. 2021]. As Tabelas 1–4 indicam que o ViT alcança maior precisão, mesmo em condições de variação de ruído e de borramento, devido à autoatenção, que melhora a representação espacial e supera CNNs [Dosovitskiy et al. 2021]. A Figura 5 mostra imagens com ruídos (Sal & Pimenta e Gaussiano), e a Figura 6, com borramento (filtros de tamanho 3x3 a 11x11).

Tabela 1. Resultados para ViT (nosso), CNN, CNN+AE, CNN+RF, CNN+SVM e AE.

Métodos	Acurácia
CNN [Silva et al. 2022a]	91.668 ± 1.528
CNN + AE [Hida et al. 2021]	91.683 ± 1.792
CNN + RF [Dong et al. 2018]	92.150 ± 1.470
CNN + SVM [Kumaresan et al. 2021]	92.876 ± 2.680
AE [Hida et al. 2021]	92.570 ± 1.178
Nosso	**95.830 ± 1.053**

5. Conclusão e Trabalhos Futuros

Este trabalho abordou a inspeção visual automática de máscaras de solda em chips GSM. A abordagem proposta alcançou alta precisão na inspeção de máscaras complexas, melhorando o processo de soldagem. Resultados experimentais confirmam sua viabilidade em aplicações reais com alta acurácia e robustez sob diferentes condições, garantindo confiabilidade. Isso reforça sua eficiência para uso industrial. Trabalhos futuros explorarão formas para melhorar a precisão e ampliar sua aplicação em outros desafios de soldagem.



(a) (b) (c)

Figura 5. 5a sem ruído, 5b com ruído Sal e Pimenta (densidade 0,02), 5c com ruído Gaussiano (densidade 0,02).

Tabela 2. Resultados com ruído Sal e Pimenta em classificação.

	Sal e Pimenta		
Densidade de ruído	**0.005**	**0.01**	**0.02**
CNN [Silva et al. 2022a]	91.118 ± 1.037	90.341 ± 1.399	89.180 ± 2.765
CNN + AE [Hida et al. 2021]	91.167 ± 1.446	90.094 ± 1.806	88.586 ± 1.992
CNN + RF [Dong et al. 2018]	92.959 ± 1.502	91.218 ± 1.808	91.020 ± 2.410
CNN + SVM [Kumaresan et al. 2021]	92.850 ± 1.232	92.305 ± 1.298	92.176 ± 2.881
AE [Hida et al. 2021]	90.440 ± 1.040	89.990 ± 1.597	89.070 ± 1.776
Nosso	**95.152 ± 1.095**	**95.007 ± 1.116**	**94.943 ± 1.237**

Tabela 3. Resultados com ruído Gaussiano em classificação.

	Gaussiano		
Densidade de ruído	**0.005**	**0.01**	**0.02**
CNN [Silva et al. 2022a]	91.431 ± 1.337	90.843 ± 1.599	90.480 ± 2.967
CNN + AE [Hida et al. 2021]	91.567 ± 1.648	90.636 ± 1.706	89.683 ± 1.892
CNN + RF [Dong et al. 2018]	92.677 ± 1.396	91.416 ± 1.440	91.105 ± 2.420
CNN + SVM [Kumaresan et al. 2021]	93.057 ± 1.131	92.405 ± 1.199	91.276 ± 2.780
AE [Hida et al. 2021]	91.540 ± 1.140	90.940 ± 1.190	89.570 ± 1.378
Nosso	**95.284 ± 1.086**	**95.011 ± 1.194**	**94.175 ± 1.430**

(a) (b) (c) (d) (e)

Figura 6. Em 6a, 6b, 6c, 6d e 6e tem filtros Gaussianos 3×3, 5×5, 7×7, 9×9 e 11×11.

Tabela 4. Resultados com imagens desfocadas, usando filtro gaussiano.

	Filtro Gaussiano Passa Baixa				
Tamanho do núcleo	**3x3**	**5x5**	**7x7**	**9x9**	**11x11**
CNN [Silva et al. 2022a]	75.876 ± 3.447	63.967 ± 4.552	61.581 ± 4.756	60.532 ± 4.973	57.312 ± 5.226
CNN + AE [Hida et al. 2021]	70.090 ± 3.788	64.679 ± 3.955	62.448 ± 4.050	58.345 ± 4.125	52.895 ± 4.330
CNN + RF [Dong et al. 2018]	83.179 ± 3.147	76.520 ± 3.360	73.433 ± 3.529	71.120 ± 3.890	71.089 ± 4.065
CNN + SVM [Kumaresan et al. 2021]	85.718 ± 2.978	84.150 ± 3.170	80.347 ± 3.677	77.368 ± 3.805	77.050 ± 3.972
AE [Hida et al. 2021]	69.322 ± 3.936	62.268 ± 4.690	60.035 ± 4.910	57.180 ± 5.440	51.806 ± 5.724
Nosso	**94.013 ± 1.505**	**93.681 ± 1.917**	**93.107 ± 2.438**	**92.117 ± 2.910**	**91.424 ± 3.602**

Referências

Al Radi, M., Li, P., Karki, H., Werghi, N., Javed, S., and Dias, J. (2023). Multi-view inspection of flare stacks operation using a vision-controlled autonomous uav. In *IECON 2023- 49th Annual Conference of the IEEE Industrial Electronics Society*, pages 1–6.

Beuth, F., Schlosser, T., Friedrich, M., and Kowerko, D. (2020). Improving automated visual fault detection by combining a biologically plausible model of visual attention with deep learning. In *IECON 2020 The 46th Annual Conf. of the IEEE Industrial Elect. Soc.*, pages 5323–5330.

Dong, X., Taylor, C. J., and Cootes, T. F. (2018). Small defect detection using convolutional neural network features and random forests. In *Proceedings of the European Conference on Computer Vision (ECCV) Workshops*, pages 0–0.

Dosovitskiy, A., Beyer, L., Kolesnikov, A., Weissenborn, D., Zhai, X., Unterthiner, T., Dehghani, M., Minderer, M., Heigold, G., Gelly, S., Uszkoreit, J., and Houlsby, N. (2021). An image is worth 16x16 words: Transformers for image recognition at scale.

Heng, P. Y. (2022). Detection of defect regions in solder mask peel off images using you-only-look-once convolutional neural network.

Hida, Y., Makariou, S., and Kobayashi, S. (2021). Smart image inspection using defect-removing autoencoder. *Proc. CIRP*, 104:559–564.

Iglesias, B. P., Otani, M., and Oliveira, F. G. (2021). Glue level estimation through automatic visual inspection in pcb manufacturing. In *ICINCO*, pages 731–738.

Kumaresan, S., Aultrin, K. J., Kumar, S., and Anand, M. D. (2021). Transfer learning with cnn for classification of weld defect. *Ieee Access*, 9:95097–95108.

Rocha, C. S., Menezes, M. A., and Oliveira, F. G. (2016). Detecção automática de microcomponentes smt ausentes em placas de circuito impresso. In *Workshop on Industry Applications (WIA) in the 29th Conference on Graphics, Patterns and Images (SIBGRAPI 2016), São José dos Campos, Sp, Brazil*, volume 1.

Silva, C. N., Ferreira, N. P., Meireles, S. S., Otani, M., da Silva, V. J., de Freitas, C. A. O., and Oliveira, F. G. (2022a). The visual inspection of solder balls in semiconductor encapsulation. In *ICINCO*, pages 750–757.

Silva, D., Lopes, A. P., Costa, D., Cabral, J., Silva, C. A., and Lopes, S. (2022b). A dataflow execution engine for automatic visual inspection of production lines. In *IECON 2022 – 48th Annual Conf. of the IEEE Industrial Elect. Society*, pages 1–6.

Swathi, G., Reddy B, R., Reddy M, A., and Kumar, R. P. (2023). Automated printed circuit board inspection: Incorporating yolov8 for efficient defect detection. In *2023 6th Int. Conf. on Recent Trends in Advance Computing (ICRTAC)*, pages 742–747.

Wang, Y.-C. Z., Huang, P.-W., and Wang, F.-C. (2023). Bubble object witness inspection equipment - a machine vision based bubble inspection device for petroleum industry. In *IECON 2023- 49th Annual Conf. of the IEEE Industrial Elect. Society*, pages 1–4.

Wu, H., Lei, R., and Peng, Y. (2022). Pcbnet: A lightweight convolutional neural network for defect inspection in surface mount technology. *IEEE Trans. on Instrumentation and Measurement*, 71:1–14.

Zhang, Q., Zhang, M., Gamanayake, C., Yuen, C., Geng, Z., Jayasekara, H., Woo, C.-w., Low, J., Liu, X., and Guan, Y. L. (2022). Deep learning based solder joint defect detection on industrial printed circuit board x-ray images. *Complex amp; Int. Systems*, 8(2):1525–1537.