Detekcija Logičnih Anomalij z Uporabo Velikih Jezikovnih Modelov

Matic Fučka[1], **Danijel Skočaj**[1]

[1]*Univerza v Ljubljani, Fakulteta za Računalništvo in Informatiko*
E-pošta: {matic.fucka, danijel.skocaj}@fri.uni-lj.si

Logical Anomaly Detection using Large Language Models

Anomaly detection is essential in industrial inspection and has recently been divided into two tasks: structural and logical anomaly detection. Structural anomaly detection focuses on visible defects such as dents or scratches, while logical anomaly detection identifies inconsistencies such as incorrect object combinations. Unlike structural anomalies, logical anomalies cannot be easily identified from a single image, as they often require an understanding of contextual relationships. We propose a new problem: Zero-shot Logical Anomaly Detection, in which only category-specific logical constraints in text form are provided at training time. The model must then determine whether an image complies with these constraints, without having seen any normal or anomalous samples. To enable this, we extend two existing datasets, MVTec LOCO and CAD-SD, with constraint annotations. We also propose a method based on Large Language Models (LLMs), prompted with chain-of-thought reasoning, to assess compliance with the given constraints. Our approach achieves AUROC scores of 69.8% on MVTec LOCO and 99.4% on CAD-SD, demonstrating the potential of LLMs in anomaly detection without visual training data.

1 Uvod

Detekcija anomalij ima ključno vlogo pri industrijskih procesih [1, 2, 3], medicinski obravnavi [4] in avtonomnimih vozilih [5]. V industrijskih procesih se to tipično loči na dve podnalogi: detekcijo strukturnih anomalij [6] in detekcijo logičnih anomalij [1, 2]. Strukturne anomalije so lokalna odstopanja videza (npr. praske), ki jih tipično detektiramo z modeliranjem običajnega izgleda. Logične anomalije pa kršijo vnaprej postavljena semantična pravila, npr. napačno število ali razporeditev delov na sliki. Te zahtevajo širše razumevanje, saj so posamezni deli povsem običajnega videza.

Najnovejši pristopi za detekcijo strukturnih anomalij brez učnih primerov (zero-shot) [7, 8, 9] temeljijo na ideji, da so si strukturne anomalije podobne med seboj pri različnih objektih. To ne velja za logične anomalije, kjer razlike pogosto najlažje in najbolje opišemo z besedilom. Iz tega sledi, da je normalnost smiselno predstaviti zgolj z besedilnimi opisi (Slika 1).

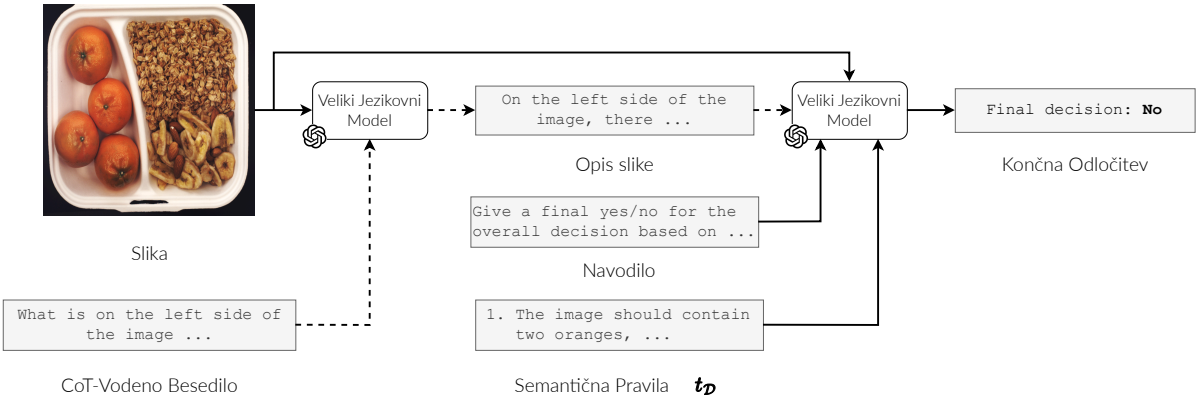
Slika 1: Največja razlika med strukturnimi in logičnimi anomalijami je v tem, katera modalnost je primernejša za opis normalnosti oziroma nenormalnosti slike. Pri strukturnih anomalijah je odstopanje pogosto najlažje opisati vizualno, medtem ko je pri logičnih anomalijah bolj smiselno tako normalnost kot abnormalnost opisati z besedilom.

Predlagamo torej nov problem: detekcija logičnih anomalij brez učnih primerov, kjer so za vsako kategorijo vnaprej podane le besedilne omejitve. Za namen tega smo razširili podatkovni množici MVTec LOCO [1] in CAD-SD [2] z opisi semantičnih pravil za vsako kategorijo. Poleg tega predstavimo protokol za vrednotenje in primerjamo več velikih jezikovnih modelov (LLM) kako se odrežejo pri tem problemu. Ker brez dodatkov ne dosegajo dobrih rezultatov, predlagamo še uporabo verižnega načina sklepanja (Chain-of-Thought, CoT) za izboljšano uspešnost.

Naši glavni prispevki v tem članku so naslednji:

- Predlagamo nov problem – detekcija logičnih anomalij brez učnih primerov in razvijemo pripadajoč protokol za vrednotenje. Na tem problemu ovrednotimo nekaj najpopularnejših velikih jezikovnih modelov.
- Uvedemo strategijo verižnega sklepanja za izboljšano delovanje LLM-jev pri predlaganemu problemu. Ta najprej izlušči opis slike, ki posledično omogoči boljšo uspešnost pri detekciji.

Eksperimente smo izvedli na dveh priljubljenih podatkovnih množicah – MVTec LOCO [1] in CAD-SD [2]. GPT-4o z predlagano CoT strategijo doseže rezultat v AUROC z vrednostima 69.8% oziroma 99.4%.



Slika 2: Slika prikazuje potek pristopa za detekcijo logičnih anomalij. V kolikor uporabljamo CoT-voden poziv najprej LLM prejme besedilo, specifično za kategorijo, da iz slike izlušči natančen opis, ki vsebuje ključne informacije za prepoznavo anomalij. Nato LLM skupaj z opisom slike in logičnimi omejitvami presodi, ali slika izpolnjuje zadane pogoje. Ta dvofazni pristop omogoča globlje razumevanje slike in bolj natančno detekcijo logičnih nepravilnosti. V kolikor ne uporabimo CoT-Voden pristop podamo LLM-ju samo sliko in sematična pravila ter se LLM odloči le na podlagi tega.

2 Sorodna Dela

Detekcija strukturnih anomalij je uveljavljeno področje v računalniškem vidu. Metode razdelimo v tri glavne skupine: rekonstrukcijske [10], diskriminativne [11, 12, 13, 14, 15] in metode na osnovi značilk [16, 17]. Rekonstrukcijske metode se učijo rekonstuirati slike brez anomalij; posledično odstopanja v rekonstrukciji nakazujejo anomalijo. Diskriminativne metode uporabljajo sintetične anomalije za učenje odločitvene meje med normalnimi in anomalnimi primeri. Metode na osnovi značilk pa uporabijo prednaučene modele in iz njih izvlečejo značilke na podlagi katerih modelirajo normalnost, npr. s spominskimi bankami.

Detekcija logičnih anomalij postaja vse pomembnejše podpodročje detekcije anomalij. Tudi tu metode delimo v tri skupine: lokalno-globalne rekonstrukcijske [18], metode, ki modelirajo globalno porazdelitev [19] in kompozicijske metode [20, 21]. Prve uporabljajo lokalne in globalne značilke za obnovo slik brez anomalij. Drugi pristop modelira porazdelitev globalnih značilk podobno kot metode na osnovi značilk za detekcijo strukturnih anomalij. Kompozicijske metode pa uporabljajo dodatne informacije, kot so kompozicijske mape (tj. semantične segmentacije brez vnaprej znanih razredov), da model vodijo v učenje pomembnih semantičnih razmerij.

Detekcija strukturnih anomalij brez učnih primerov temelji na predpostavki, da so vizualne anomalije podobne med različnimi objekti. Pristopi navadno uporabljajo zunanje podatkovne zbirke (npr. učenje modela na MVTec AD [6] in testiranje na VisA [22]) in velike vizualne modele, kot sta CLIP [23] in SAM [24]. Metode na podlagi modela CLIP [8, 7] uporabljajo ročno pripravljena besedila za detekcijo odstopanj; nekatere dodajo tudi učljive besedilne člene. Metode na podlagi modela SAM [25, 26] prilagodijo SAM za detekcijo anomalnih regij in izboljšanje segmentacije. Čeprav te metode dobro delujejo pri strukturnih anomalijah, niso učinkovite pri logičnih, saj brez učnih primerov ne morejo razumeti logičnih pravil.

Detekcija logičnih anomalij z Velikimi Jezikovnimi Modeli je v zadnjem letu pridobila veliko pozornosti z uvedbo LogicAD [27], kjer se LLM uporablja za ekstrakcijo opisa slike, medtem ko uporabijo klasične metode za dokončno detekcijo anomalije. Kasnejši pristopi [28, 29] izboljšajo besedilna sporočila ali uporabijo GRPO [30] za izboljšanje možnosti sklepanja LLM-ja. Kljub temu se ti modeli še vedno zanesejo večinoma na modele, ki temeljijo zgolj na slikah, kot je GroundingDINO [31], kar pomeni, da se večinoma zanašajo na vizualne namige in nimajo na voljo visoko-nivojskih informacij potrebnih za detekcijo logičnih anomalij. Za razliko od prejšnih pristopov mi predlagamo metodo, ki neposredno uporablja sposobnost sklepanja pri LLM-jih brez da bi se zanašali na modele, ki uporabljajo le vizualne informacije.

3 Detekcija logičnih anomalij brez učnih primerov

Za nadaljnji razvoj metod za detekcijo logičnih anomalij brez učnih primerov bomo najprej definirali postavitev problema in mere uspešnosti.

3.1 Definicija problema

Naj bo $\mathcal{D} = I_1, I_2, ..., I_N$ testna množica N slik, kjer nekatere vsebujejo logične anomalije. Poleg teh slik je podano še besedilo $t_{\mathcal{D}}$, ki v naravnem jeziku določa semantična pravila. Cilj je razviti detektor anomalij $\tau : I_i \rightarrow 0, 1$, ki zna ločiti slike z logičnimi anomalijami od tistih brez njih.

3.2 Mere uspešnosti

Uspešnost detekcije logičnih anomalij brez učnih primerov ocenjujemo z dvema uveljavljenima merama: površino pod ROC krivuljo (AUROC) in maksimalnim F_1-rezultatom (F_1-max). Ti dve meri smo izbrali, ker vsaka oceni drug aspekt modela. AUROC meri, kako dobro model detekira

Metoda	Breakfast Box	Juice Bottle	Pushpins	Screw Bag	Splicing Connectors	*Povprečje*
AnomalyCLIP [7]	62.7	52.5	50.9	55.4	59.5	56.3
AdaCLIP [8]	51.3	67.0	48.4	51.2	61.3	55.9
GPT-4o-mini	68.9	63.7	56.4	50.2	51.9	58.1
GPT-4o-mini *CoT-Vodeno*	73.6	60.3	62.2	52.1	65.3	62.7 (+4.6)
GPT-4.1	61.3	78.9	50.7	49.4	60.3	59.9
GPT-4.1 *CoT-Vodeno*	78.8	79.7	58.2	56.2	58.7	66.4 (+6.5)
GPT-4o	71.1	76.8	50.2	55.1	69.6	64.7
GPT-4o *CoT-Vodeno*	78.0	75.3	59.0	62.6	74.0	69.8 (+5.1)

Tabela 1: Rezultati detekcije anomalij (AUROC) na podatkovni množici MVTec LOCO [1].

Metoda	Breakfast Box	Juice Bottle	Pushpins	Screw Bag	Splicing Connectors	*Povprečje*
AnomalyCLIP [7]	62.3	67.3	56.2	53.6	61.8	60.3
AdaCLIP [8]	61.9	65.7	57.1	59.2	64.7	61.7
GPT-4o-mini	65.5	57.0	51.5	21.8	7.1	40.6
GPT-4o-mini *CoT-Vodeno*	71.1	76.6	51.5	50.9	64.9	63.0 (+22.4)
GPT-4.1	66.9	80.7	57.2	60.2	64.3	66.7
GPT-4.1 *CoT-Vodeno*	78.1	83.2	58.8	64.1	60.5	69.2 (+2.5)
GPT-4o	72.1	69.7	56.8	50.0	58.2	61.4
GPT-4o *CoT-Vodeno*	76.3	78.8	58.8	57.1	67.4	67.7 (+6.3)

Tabela 2: Rezultati detekcije anomalij (F_1-max) na podatkovni množici MVTec LOCO [1].

Metoda	CAD-SD	
	AUROC	F_1-max
AnomalyClip [7]	52.7	61.4
AdaCLIP [8]	52.9	61.5
GPT-4o-mini	64.3	44.4
GPT-4o-mini *CoT-Vodeno*	86.9(+22.5)	75.3(+30.9)
GPT-4.1	86.0	74.0
GPT-4.1 *CoT-Vodeno*	95.7(+9.7)	90.3(+16.3)
GPT-4o	85.7	73.7
GPT-4o *CoT-Vodeno*	99.4(+13.7)	99.4(+25.7)

Tabela 3: Rezultati detekcije anomalij na podatkovni množici CAD-SD [2].

anomalije nad slikami brez anomalij pri vseh pragih. F_1-max zajema najboljše ravnotežje med natančnostjo (precision) in priklicem (recall) pri enem pragu. Uporaben je za oceno uspešnosti odločanja, ko je pomembna tudi pravilna klasifikacija med anomalnimi in neanomalnimi slikami.

4 Detekcija logičnih anomalij z LLM-ji

Poleg predloga novega problema predstavimo tudi dva različna načina uporabe LLM-jev za rešitev le tega. Prvi, neposredni pristop, temelji na neposredni oceni, ali slika ustreza logičnim omejitvam zgolj na podlagi vizualnih značilnosti. Drugi, CoT-voden pristop, pa najprej iz slike izlušči podroben opis in nato LLM oceni logično skladnost glede na vizualne in besedilne informacije skupaj.

4.1 Neposredni pristop

Za osnovno oceno sposobnosti LLM-jev pri analizi slike uporabimo preprost pristop: modelu podamo sliko I in semantična pravila $t_{\mathcal{D}}$ ter ga vprašamo, ali slika ugaja tem pravilom. Rezultat je ena sama binarna odločitev (1/0) za vsako sliko.

4.2 CoT-vodena detekcija anomalij

Novejše raziskave [32] kažejo, da kombinacija slike in njenega opisa bistveno izboljša rezultate. Na tej osnovi razvijemo dvofazni pristop. V prvi fazi LLM z uporabo specifičnega besedila za vsako kategorijo ustvari opis slike, ki vsebuje le informacije, pomembne za detekcijo logičnih anomalij. V drugi fazi model prejme sliko, opis in omejitve ter presodi, ali so te zadovoljene. Tudi tu mora model vrniti končno odločitev 1/0 za vsako sliko. Primer takega pozivanja je prikazan na Sliki 2.

5 Eksperimenti

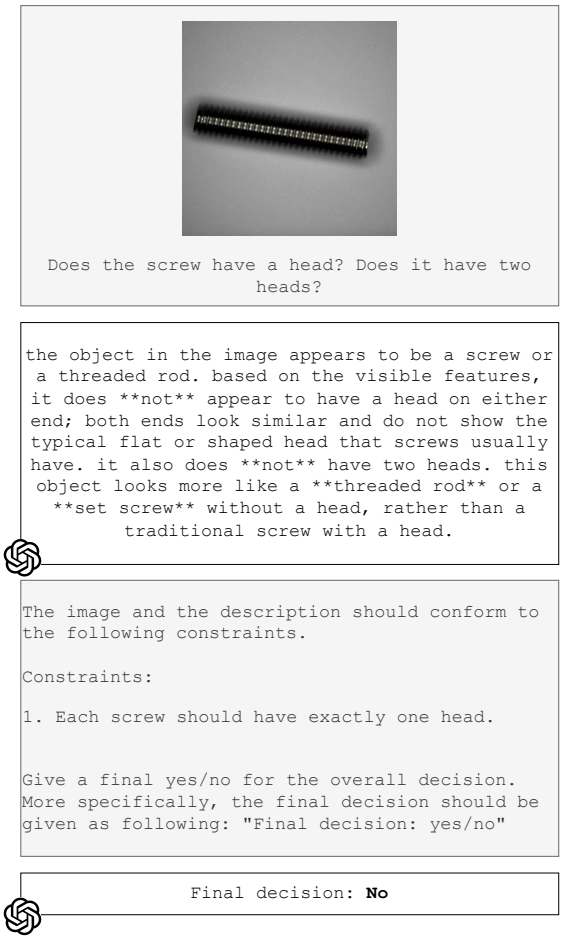
5.1 Podatkovne množice

Eksperimente izvajamo na dveh standardnih podatkovnih množicah za detekcijo logičnih anomalij: MVTec LOCO [1] in CAD-SD [2]. MVTec LOCO vsebuje 3,644 slik v petih kategorijah z oznakami na ravni slik in pikslov, medtem ko CAD-SD vsebuje 776 slik ene same kategorije in ponuja le oznake na ravni slike. V vrednotenju uporabimo le podmnožice brez anomalij in z logičnimi anomalijami; pri CAD-SD to ustreza podmnožicama Overcoupling in Lacking. Ker nobena od zbirk ne vsebuje besedilnih opisov omejitev, smo jih sami pripravili za vsako kategorijo.

5.2 Rezultati

Za oceno obeh pristopov smo uporabili več trenutno najuspešnejših LLM-jev: GPT-4o-mini [33], GPT-4o [33] in GPT-4.1 [34]. Primerjali smo neposredni pristop z našim CoT-vodenim pozivanjem, da ocenimo vpliv dodatnega sklepanja.

Rezultati na MVTec LOCO so prikazani v Tabelah 1 in 2. Naš pristop izboljša rezultate pri vseh modelih in po



Slika 3: Kvalitativni prikaz delovanja CoT-Vodenega načina pozivanja na primeru iz CAD-SD [2]. Metoda najprej pridobi natančen opis slike in se nato na podlagi slike in omejitev opredeli, če je s sliko kaj narobe. Pozivi s strani uporabnika so prikazani v sivi barvi, odgovori s strani LLM-ja pa v beli barvi.

obeh merah uspešnosti (za $+4,6$ odstotnih točk (o. t.) pri GPT-4o-mini, $+6,5$ o. t. pri GPT-4.1 glede na AUROC). To potrjuje, da opis slike pomaga pri natančnejši oceni skladnosti z omejitvami. Najvišji AUROC doseže GPT-4o, medtem ko GPT-4.1 doseže najvišji F_1-max — kar kaže na razlike v močnih straneh posameznih modelov. Kvalitativni primer delovanja je prikazan tudi na Sliki 3.

Rezultati na CAD-SD (Tabela 3 in Slike 3) pokažejo še večje izboljšave z izboljšanim pozivanjem: $+22,5$, $+9,7$ in $+13,5$ o. t. za GPT-4o-mini, GPT-4.1 in GPT-4o glede na AUROC. Podobno velja tudi za F_1-max. S tako visokimi rezultati naša metoda prekaša tudi znane modele, kot sta PatchCore [16] in DRÆM [12]. To potrjuje, da so LLM-ji primerna izbira za detekcijo logičnih anomalij brez predhodnega učenja.

Poleg tega naš pristop prekaša vse obstoječe metode za detekcijo anomalij brez učnih primerov, kar je pričakovano, saj so bile te metode zasnovane za strukturne anomalije. Rezultati kažejo, da je za uspešno detekcijo logičnih anomalij nujno poglobljeno razumevanje vsebine slike — kar CLIP [23], uporabljen kot osnova v večini primerjalnih metod, ne ponuja v zadostni meri.

6 Zaključek

V tem članku smo predstavili nov problem: detekcijo logičnih anomalij brez učnih primerov in shemo vrednotenja za ta problem. V tem problemu so med učenjem podane le logične omejitve za posamezno kategorijo, model pa mora normalnost sklepati zgolj iz teh besedilnih opisov. Za izhodišče bodočemu raziskovanju smo ovrednotili tudi uspešnost nekaj priljubljenih velikih jezikovnih modelov in predlagali izboljšavo pozivanja s pomočjo verižnega načina sklepanja (CoT). Z izboljšanim načinom pozivanja dosežemo AUROC 69.8% na MVTec LOCO [1] in 99.4% na CAD-SD [2]. V prihodnosti bi bilo to mogoče izboljšati s podajanjem dodatnih informacij, pridobljenih iz Vizualnih Temeljnih Modelov (ang. Vision Foundation Models), v veliki jezikovni model.

7 Zahvala

Raziskave so bile delno podprte s projektom MUXAD (J2-60055), raziskovalnim programom P2-0214 ter superračunalniškim omrežjem SLING (ARNES, EuroHPC Vega - IZUM).

Literatura

[1] P. Bergmann, K. Batzner, M. Fauser, D. Sattlegger, and C. Steger, "Beyond Dents and Scratches: Logical Constraints in Unsupervised Anomaly Detection and Localization," *International Journal of Computer Vision*, vol. 130, no. 4, pp. 947–969, 2022.

[2] K. Ishida, Y. Takena, Y. Nota, R. Mochizuki, I. Matsumura, and G. Ohashi, "Sa-PatchCore: Anomaly Detection in Dataset with Co-Occurrence Relationships Using Self-Attention," *IEEE Access*, vol. 11, pp. 3232–3240, 2023.

[3] J. Božič, M. Fučka, V. Zavrtanik, and D. Skočaj, "Robustness of unsupervised methods for image surface-anomaly detection," *Pattern Analysis and Applications*, vol. 28, p. 99, May 2025.

[4] B. H. Menze, A. Jakab, S. Bauer, J. Kalpathy-Cramer, K. Farahani, J. Kirby, Y. Burren, N. Porz, J. Slotboom, R. Wiest, *et al.*, "The multimodal brain tumor image segmentation benchmark (brats)," *IEEE transactions on medical imaging*, vol. 34, no. 10, pp. 1993–2024, 2014.

[5] H. Blum, P.-E. Sarlin, J. Nieto, R. Siegwart, and C. Cadena, "The fishyscapes benchmark: Measuring blind spots in semantic segmentation," *International Journal of Computer Vision*, vol. 129, no. 11, pp. 3119–3135, 2021.

[6] P. Bergmann, K. Batzner, M. Fauser, D. Sattlegger, and C. Steger, "The MVTec Anomaly Detection Dataset: A Comprehensive Real-World Dataset for Unsupervised Anomaly Detection," *International Journal of Computer Vision*, vol. 129, no. 4, pp. 1038–1059, 2021.

[7] Q. Zhou, G. Pang, Y. Tian, S. He, and J. Chen, "AnomalyCLIP: Object-agnostic Prompt Learning for Zero-shot Anomaly Detection," in *The Twelfth International Conference on Learning Representations, ICLR 2024, Vienna, Austria, May 7-11, 2024*, 2024.

[8] Y. Cao, J. Zhang, L. Frittoli, Y. Cheng, W. Shen, and G. Boracchi, "AdaCLIP: Adapting CLIP with Hybrid Learnable Prompts for Zero-Shot Anomaly Detection," in *European Conference on Computer Vision*, pp. 55–72, Springer, 2024.

[9] J. Jeong, Y. Zou, T. Kim, D. Zhang, A. Ravichandran, and O. Dabeer, "WinCLIP: Zero-/Few-Shot Anomaly Classification and Segmentation," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 19606–19616, 2023.

[10] V. Zavrtanik, M. Kristan, and D. Skočaj, "Reconstruction by inpainting for visual anomaly detection," *Pattern Recognition*, vol. 112, p. 107706, 2021.

[11] M. Fučka, V. Zavrtanik, and D. Skočaj, "TransFusion – A Transparency-Based Diffusion Model for Anomaly Detection," in *European conference on computer vision*, pp. 91–108, Springer, 2025.

[12] V. Zavrtanik, M. Kristan, and D. Skočaj, "DRAEM-A discriminatively trained reconstruction embedding for surface anomaly detection," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pp. 8330–8339, 2021.

[13] V. Zavrtanik, M. Kristan, and D. Skočaj, "DSR–A dual subspace re-projection network for surface anomaly detection," in *European Conference on Computer Vision*, pp. 539–554, Springer, 2022.

[14] B. Rolih, M. Fučka, and D. Skočaj, "SuperSimpleNet: Unifying Unsupervised and Supervised Learning for Fast and Reliable Surface Defect Detection," in *International Conference on Pattern Recognition*, 2024.

[15] B. Rolih, M. Fučka, and D. Skočaj, "No Label Left Behind: A Unified Surface Defect Detection model for all Supervision Regimes," *Journal of Intelligent Manufacturing*, 2025.

[16] K. Roth, L. Pemula, J. Zepeda, B. Schölkopf, T. Brox, and P. Gehler, "Towards Total Recall in Industrial Anomaly Detection," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 14318–14328, 2022.

[17] H. Deng and X. Li, "Anomaly Detection via Reverse Distillation From One-Class Embedding," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 9737–9746, June 2022.

[18] K. Batzner, L. Heckler, and R. König, "EfficientAD: Accurate Visual Anomaly Detection at Millisecond-Level Latencies," in *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, pp. 128–138, 2024.

[19] N. Cohen, I. Tzachor, and Y. Hoshen, "Set Features for Anomaly Detection," *arXiv preprint arXiv:2311.14773*, 2023.

[20] T. Liu, B. Li, X. Du, B. Jiang, X. Jin, L. Jin, and Z. Zhao, "Component-aware anomaly detection framework for adjustable and logical industrial visual inspection," *Advanced Engineering Informatics*, vol. 58, p. 102161, 2023.

[21] M. Fučka, V. Zavrtanik, and D. Skočaj, "SALAD – Semantics-Aware Logical Anomaly Detection," in *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, October 2025.

[22] Y. Zou, J. Jeong, L. Pemula, D. Zhang, and O. Dabeer, "SPot-the-Difference Self-supervised Pre-training for Anomaly Detection and Segmentation," in *Computer Vision–ECCV 2022: 17th European Conference, Tel Aviv, Israel, October 23–27, 2022, Proceedings, Part XXX*, pp. 392–408, Springer, 2022.

[23] A. Radford, J. W. Kim, C. Hallacy, A. Ramesh, G. Goh, S. Agarwal, G. Sastry, A. Askell, P. Mishkin, J. Clark, *et al.*, "Learning Transferable Visual Models from Natural Language Supervision," in *International conference on machine learning*, pp. 8748–8763, PMLR, 2021.

[24] A. Kirillov, E. Mintun, N. Ravi, H. Mao, C. Rolland, L. Gustafson, T. Xiao, S. Whitehead, A. C. Berg, W.-Y. Lo, *et al.*, "Segment anything," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pp. 4015–4026, 2023.

[25] Y. Cao, X. Xu, Y. Cheng, C. Sun, Z. Du, L. Gao, and W. Shen, "Personalizing vision-language models with hybrid prompts for zero-shot anomaly detection," *IEEE Transactions on Cybernetics*, vol. 55, no. 4, pp. 1917–1929, 2025.

[26] S. Li, J. Cao, P. Ye, Y. Ding, C. Tu, and T. Chen, "Clipsam: Clip and sam collaboration for zero-shot anomaly segmentation," *Neurocomputing*, vol. 618, p. 129122, 2025.

[27] E. Jin, Q. Feng, Y. Mou, S. Decker, G. Lakemeyer, O. Simons, and J. Stegmaier, "LogicAD: Explainable Anomaly Detection via VLM-based Text Feature Extraction," 2025.

[28] W. Li, G. Chu, J. Chen, G.-S. Xie, C. Shan, and F. Zhao, "Lad-reasoner: Tiny multimodal models are good reasoners for logical anomaly detection," *arXiv preprint arXiv:2504.12749*, 2025.

[29] Y. Kwon, D. Moon, Y. Oh, and H. Yoon, "Logicqa: Logical anomaly detection with vision language model generated questions," *arXiv preprint arXiv:2503.20252*, 2025.

[30] Z. Shao, P. Wang, Q. Zhu, R. Xu, J. Song, X. Bi, H. Zhang, M. Zhang, Y. Li, Y. Wu, *et al.*, "DeepSeekMath: Pushing the limits of mathematical reasoning in open language models," *arXiv preprint arXiv:2402.03300*, 2024.

[31] S. Liu, Z. Zeng, T. Ren, F. Li, H. Zhang, J. Yang, Q. Jiang, C. Li, J. Yang, H. Su, *et al.*, "Grounding DINO: Marrying DINO with Grounded Pre-Training for Open-Set Object Detection," in *European Conference on Computer Vision*, pp. 38–55, Springer, 2025.

[32] Z. Zhang, A. Zhang, M. Li, H. Zhao, G. Karypis, and A. Smola, "Multimodal chain-of-thought reasoning in language models," *arXiv preprint arXiv:2302.00923*, 2023.

[33] A. Hurst, A. Lerer, A. P. Goucher, A. Perelman, A. Ramesh, A. Clark, A. Ostrow, A. Welihinda, A. Hayes, A. Radford, *et al.*, "GPT-4o system card," *arXiv preprint arXiv:2410.21276*, 2024.

[34] OpenAI, "GPT-4 technical report," 2024.