

# Detekcija površinskih napak na oblačilih za reciklažo z uporabo nadzorovanih metod globokega učenja

Domen Tabernik, Danijel Skočaj

Univerza v Ljubljani, Fakulteta za računalništvo in informatiko

E-pošta: {domen.tabernik,danijel.skocaj}@fri.uni-lj.si

## Supervised Deep Learning for Surface Defect Detection in Recyclable Garments

*Efficient sorting of used garments is essential for textile recycling in the circular economy. Surface defect detection, such as identifying stains or tears, enables automated classification of items for reuse or recycling. In this paper, we focus on the problem of detecting surface defects on second-hand clothing using supervised deep learning methods. We present an analysis of our two previously proposed general-purpose surface defect detection models (SegDecNet and SuperSimpleNet) along with four modern backbone image classification architectures (ConvNeXt, ViT, Swin, and DINO). For evaluation, we curate a tailored binary classification dataset derived from the real-world garment dataset, including over 12 000 annotated clothing images. Our results show that SuperSimpleNet significantly outperforms other methods, achieving an average precision of 72%, while highlighting the inherent challenges of this task due to garment variability and subtle or occluded defects.*

### 1 Uvod

V krožnem gospodarstvu ima reciklaža oblačil ključno vlogo pri zagotavljanju trajnostnega razvoja ter zmanjšanju okoljskih vplivov. Ključen korak v tem procesu je natančno sortiranje oblačil glede na njihovo stanje. Detekcija površinskih napak, kot so madeži, raztrganine ali luknje, omogoča razločevanje med oblačili, ki so še primerna za ponovno uporabo, in tistimi, ki jih je treba predelati ali zavreči. Zaradi ogromnega števila kosov, ki prihajajo v centre za reciklažo, je ročno pregledovanje neučinkovito in subjektivno, zato je avtomatizacija teh postopkov smiselna in nujna.

Globoko učenje in računalniški vid sta se v zadnjem desetletju uveljavila kot učinkovita pristopa za detekcijo površinskih napak na izdelkih. Algoritmi, pogosto razviti za industrijske predmete z doslednim videzom in pod nadzorovanimi pogoji zajema, so sposobni doseči visoke natančnosti [11]. Pri oblačilih pa je situacija bistveno zahtevnejša – kosi prihajajo v izredno raznolikih oblikah, barvah, vzorcih, materialih in stopnjah obrabe. Poleg tega niso rigidni kot običajni industrijski izdelki, kar privede do nagubanosti, delnega prekrivanja ali deformiranosti. V literaturi sicer obstajajo raziskave, ki se ukvarjajo z detekcijo napak na tekstilu, vendar so te večinoma osredo-

točene na detekcijo napak na tkaninah med proizvodnjo, kjer je tkanina vpeta v proizvodnje enote ter tako brez gub, sočasno pa je enakomerno osvetljena ter brez motenj iz ozadja [14].

V tem članku se osredotočamo na problem detekcije površinskih napak na končnih oblačilih, kot jih srečujemo v realnih postopkih sortiranja za reciklažo. Z uporabo nadzorovanih metod globokega učenja izvedemo primerjalno analizo metod za detekcijo napak na ravni slike. Za analizo smo izbrali dve metodi za detekcijo površinskih napak, ki smo ju v naših preteklih delih že predlagali za uporabo na splošnih rigidnih industrijskih izdelkih: SegDecNet [12, 1] ter SuperSimpleNet [11]. Metodi sedaj preverimo za delovanje na deformabilnih neindustrijskih izdelkih kot so oblačila. Dodatno smo preverili tudi delovanje štirih sodobnih hrbtničnih arhitektur globokega učenja za klasifikacijo slik: ConvNeXt [9], ViT [3], Swin [7] ter DINO [4]. Za namen analize smo pripravili prilagojeno množico slik, sestavljeno iz 31 000 slik velike javno dostopne baze rabljenih oblačil iz realnega industrijskega primera reciklaže [10], ki jo nato uporabimo za učenje ter vrednotenje omenjenih metod. Primerjalna analiza prikaže uspešnost metode SuperSimpleNet v primerjavi z drugimi metodami, dodatno pa osvetli tudi samo kompleksnost problema.

### 2 Sorodna dela

Detekcija površinskih napak na tekstilu se v literaturi primarno pojavlja v povezavi z analizo tkanin na slikah, ki zajamejo samo določeno tkanino povsem brez gub, v odlični osvetlitvi ter v visoko resolucijski sliki same napake. Prilagojeno za te probleme so bile predlagane različne globoke arhitekture, kot je na primer nadgrajeni RefineDet z robno usmerjenimi bloki [13] ali nadgrajeni YOLOv11 z dodatno prostorsko piramido, SPFFNet [14].

Na drugi strani se za splošno detekcijo površinskih napak pogosto uporabljajo metode kot je SegDecNet [12] ali SuperSimpleNet [11]. Metoda SegDecNet [12] omogoča detekcijo anomalij z učenjem dodatnih segmentacij, bodisi v popolnoma nadzorovanem načinu [12], bodisi v kombinaciji z mešanim nadzorom [1]. Za učenje v mešanem nadzoru je bila nedavno predstavljena tudi metoda SuperSimpleNet [11], ki temelji na nenadzorovanem pristopu učenja v kombinaciji z učenjem segmentacijske maske, s čimer doseže vrhunski rezultat na vrsti klasič-



Slika 1: Primeri poškodovanih oblačil iz podatkovne množice *Second-Hand Fashion* [10] ter ustreznimi segmentacijskimi oznakami.

nih podatkovnih množicah za detekcijo površinskih napak. Navkljub vrhunskim rezultatom, pa delovanje splošnih metod do sedaj ni bilo raziskano za problem detekcije napak na oblačilih.

### 3 Definicija problema in podatkovna množica

V tem članku se osredotočimo na problem detektiranja površinskih napak na oblačilih. Za površinske napake definiramo vse poškodbe na oblekah zaradi česar oblačilo ni več primerno za nadaljnjo uporabo. Poškodbe zajemajo razne odrgnine, raztrganine ter luknje kot tudi površinske madeže. Problem definiramo kot binarna klasifikacija slike, kjer za oblačilo na sliki napovemo prisotnost ali odsotnost poškodbe. Za učenje in vrednotenje metod globokega učenja na tem problemu definiramo podatkovno množico, ki temelji na slikah iz zbirke *Second-Hand Fashion* [10].

**Second-Hand Fashion:** Podatkovna množica *Second-Hand Fashion* [10] obsega 31 638 kosov oblačil zajetih med postopkom reciklaže v Švedskem podjetju Wargön Innovation AB. Na zajemni postaji so oblačila razgrnjena ter posneta od spredaj in od zadaj z ločljivostjo  $1920 \times 1080$  ali  $1280 \times 720$  pikslov. Oblačila so bila nato ročno označena z vrsto označb, kot so vrsta in barva oblačil, znamka, material, cena, ter podrobne informacije o stanju ter tipu reciklaže (npr., za izvoz, za popravilo, za reciklažo, za ponovno uporabo, itd.). Informacije o stanju zajemajo oceno stanja od 1 do 5 (najslabše do najboljše), oceno stanja o muckanju tkanine od 1 do 5, prisotnost (večja/majhna) luknje, madežev in vonju, ter kratek opis o potencialni poškodbi obleke. Na podlagi zajetih oblačil so avtorji pripravili učno (28 247) in testno množico slik (3389), ki pa vsebujeta le slike posnete od spredaj. Tem slikam je bilo še dodatno odstranjeno ozadje s segmentacijsko metodo BiRefNet [15], ki pa ni bila vedno popolna ter je v nekaterih primerih ostal viden tudi šiviljski trak v ozadju. Pripravljeno učno in testno množico obdelanih

sprednjih slik so avtorji naredili dostopno preko portala Hugging Face<sup>1</sup>.

**Končna zbirka slik za učenje in vrednotenje:** Pripravljene učne in testne sprednje slike iz zbirke *Second-Hand Fashion* smo dodatno razdelili za problem binarne detekcije poškodb na oblačilih (prisotnost/odsotnost poškodovane obleke na sliki). Za binarni problem smo v osnovi določili spodnje kriterije:

**poškodovano oblačilo** kakršenkoli opis v oznakah poškodbe, luknje ali madežev (razen opisa *No* ali *None*).

**ne-poškodovano oblačilo** če ne ustreza definiciji poškodovanega oblačila, nima opisa vonja ter ima oceno stanja in muckanja večje ali enako 4.

Slike oblačil, ki ne ustrezajo zgornjim kriterijem smo zavrgli, nekatere preostale pa še dodatno ročno verificirali za prisotnost poškodbe. Dodatna verifikacija je bila izvedena primarno na podmnožici slik s potencialnimi poškodbami glede na zgornji kriterij poškodovanosti, saj originalni atributi poškodb pogosto niso odražali stanja na sliki. Po ročni verifikaciji smo obdržali slike s potrjeno opazno poškodbo (pozitivni primeri), izmed tega 258 učnih ter 98 testnih slik. Ker na velikem deležu slik nismo uspeli identificirati poškodbe ocenjujemo, da so bile poškodbe bodisi na drugi strani oblačila bodisi je bila poškodba premajhna, da bi bila opazna na sliki. Slike na katerih nismo potrdili poškodbe, vendar so bile označene kot poškodovane, smo povsem izločili iz končne množice. Na vseh slikah poškodovanih oblek smo poškodbo tudi ročno segmentirali. Primeri slik s poškodovanimi oblačili ter ročnimi segmentacijami za učenje so prikazani na Sliki 1. Množica ne-poškodovanih oblačil (negativni primeri) obsega 10 943 učnih ter 939 testnih slik, pri čemer smo ročno verificirali le testne slike, saj so bili atributi nepoškodovanih oblačil bolj zanesljivi.

<sup>1</sup><https://huggingface.co/datasets/fnauman/fashion-second-hand-front-only-rgb>

Metoda	Seg.	Utež. krit. funk.	AP [%]	AUC [%]	Točnost [%]	F1 [%]
SuperSimpleNet [11]	✓		71,7	91,6	94,3	69,6
ViT-B [3]			61,6	87,6	93,8	59,5
Swin-B [7]			57,8	84,9	93,1	56,6
SuperSimpleNet [11]			57,4	87,1	89,6	53,7
Swin-B v2 [6]			52,8	84,2	92,1	52,3
ConvNeXt-B [9]			51,9	81,5	90,2	49,0
ViTMAE-B [5]			45,4	77,5	92,4	50,9
DINOv2-B z registri [2]			42,7	77,2	88,6	45,4
SegDecNet [12, 1]	✓	✓	38,8	76,2	90,0	38,8
SegDecNet [12, 1]	✓		34,7	73,1	85,5	38,5
SegDecNet [12, 1]			31,5	72,5	90,1	36,8
DINOv2-B [4]			24,9	70,1	82,5	35,6

Tabela 1: Rezultati različnih arhitektur na pripravljene podatkovni množici *Second-Hand Fashion*.

## 4 Metode

Za analizo smo se osredotočili na sodobne modele globokih nevronske mreže za detekcijo napake na nivoju slike oz. za klasifikacijo prisotnosti/odsotnosti napake v sliki. V ta namen smo ovrednotili dve metodi za detekcijo splošnih površinskih napak: SegDecNet [12] ter SuperSimpleNet [11]. Dodatno pa smo ovrednotili štiri vrste sodobnih hrbtničnih globokih nevronske mreže: ConvNeXt [9], ViT [3], Swin [7] ter DINO [4]. V tem razdelku podamo hiter pregled preizkušenih pristopov, ki jih uporabimo za detekcijo površinskih napak na oblačilih.

### 4.1 Splošne metode za detekcijo površinskih napak

**SegDecNet:** Metoda SegDecNet [12] je ena prvih metod globokega učenja za detekcijo površinskih napak na industrijskih izdelkih. Temelji na dvo-nivojski arhitekturi, ki združuje učenje segmentacije in klasifikacije slike, pri čemer postopoma utežuje segmentacijsko (L1) in klasifikacijsko (binarna križna entropija) funkcijo izgube.

Pri neuravnoveženih podatkih uvaja podvzorčenje negativnih primerov, da ohrani uravnoveženo razmerje med pozitivnimi in negativnimi vzorci ter enakomerno vzorčenje skozi učne epohe. Za učenje običajno potrebuje segmentacijsko masko, metoda pa z ustreznim uteževanjem omogoča tudi mešan nadzor, kjer segmentacijska maska ni na voljo za vsak pozitiven primer [1].

**SuperSimpleNet** Metoda SuperSimpleNet [11] nadgrajuje SimpleNet [8] za nenadzorovano učenje videza normalnosti na osnovi šumnih značilk iz mreže WideResNet-50. Uvede izboljšano arhitekturo, sintetično generiranje anomalij ter možnost vključitve pozitivnih primerov (slik z napakami), kar omogoča učenje tako v nenadzorovanem kot tudi popolnoma nadzorovanem načinu s segmentacijskimi oznakami. Podobno kot SegDecNet hkrati uči segmentacijo (L1) in klasifikacijo (fokalna izguba) ter uporablja podvzorčenje negativnih primerov za uravnoveženo učenje.

### 4.2 Sodobne hrbtnične arhitekture za klasifikacijo

Kot referenčne metode smo preizkusili sodobne hrbtnične arhitekture z binarnim klasifikatorjem nad predna-

učnimi značilkami, ter nato učene z binarno križno entropijo za detekcijo napak na oblačilih. Pri konvolucijskih modelih smo uporabili globalno povprečenje značilk pred klasifikacijo, vse metode pa vključujejo podvzorčenje negativnih primerov za uravnoveženo učenje, skladno s pristopi SegDecNet in SuperSimpleNet. Za primerjavo smo naučili in preizkusili sledeče arhitekture: ConvNeXt [9], Vision Transformer (ViT [3] in ViTMAE [5] naučen samo-nadzorovano), Swin Transformer (v1 [7] in v2 [6]) ter DINO (DINOv2 [4] in DINOv2 z registri [2]).

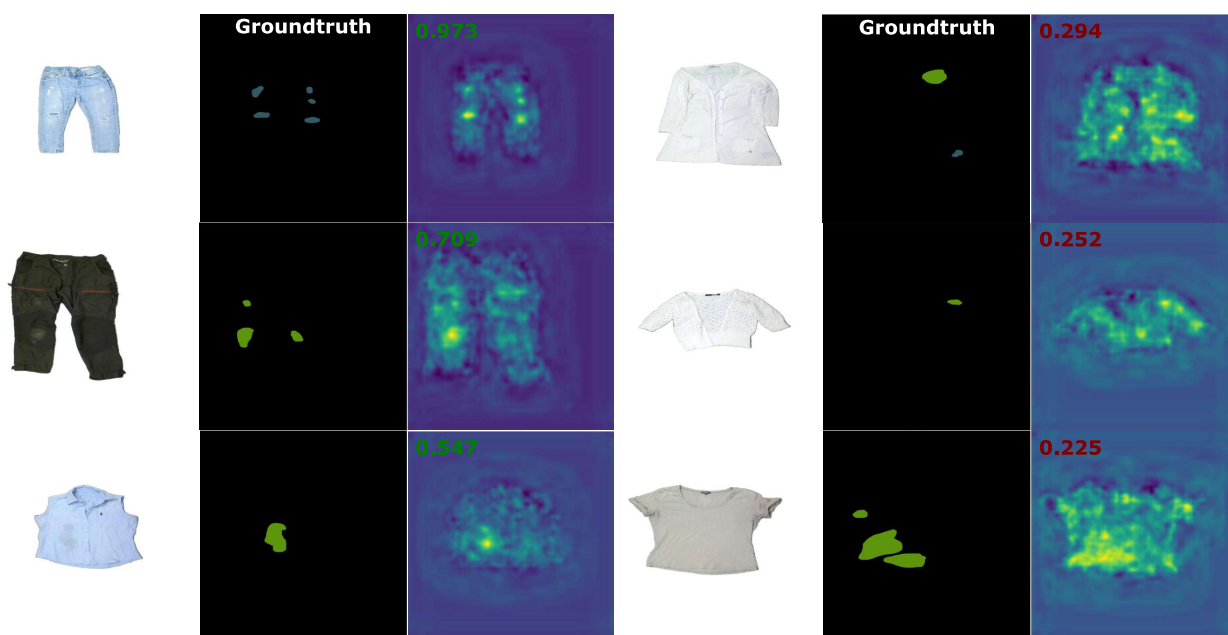
## 5 Eksperimentalni rezultati

Izbrane metode smo naučili in ocenili za problem detektiranja poškodovanih oblačil na ročno pripravljene podmnožici slik *Second-Hand Fashion* s skupno 12 238 slik oblačil.

### 5.1 Postopki učenja

Pri vseh modelih smo vhodno sliko normalizirali po ImageNet standardu ter jo zmanjšali na  $640 \times 640$  pikslov, medtem ko se je pri sodobnih hrbtničnih arhitekturah slika še dodatno zmanjšala na velikost uporabljeno pri originalnem učenju arhitekture. Pri modelih, ki uporabljajo hrbtnične arhitekture (vsi razen SegDecNet) smo uporabili tudi prednaučene uteži. Vse modele smo nato učili 500 epoh, pri čemer je ena epoha zajemala 238 normalnih slik ter 238 slik z napako oz. poškodbo na oblačilu. Pri vseh smo tudi zagotovili enakomerno uporabo negativnih primerov med celotnim učenjem.

Pri vseh modelih smo uporabili velikost paketa 48 slik ter klasično optimizacijo stohastičnega gradientnega spusta brez zmanjševanja uteži (ang. *Stochastic Gradient Descent*). Za vse modele smo preizkusili različne stopnje učenja (0,1, 0,05, 0,01, 0,001 ter 0,0001), ter jih ovrednotili na vsakih 40 epoh. Na koncu poročamo najboljši model glede na povprečno natančnost na testni množici. Za SegDecNet ter SuperSimpleNet smo dodatno preizkusili tudi model brez učenja segmentacije, ter za SegDecNet tudi model brez uteževanja funkcije izgube med segmentacijo in klasifikacijo.



Slika 2: Primeri pravilnih (levo) in zgrešenih (desno) detekcij z metodo SuperSimpleNet. Zgoraj levo je prikazana napovedana verjetnost za poškodbo oblačila (v zelenem pravilna, v rdečem napačna).

## 5.2 Rezultati

Rezultati so poročani v Tabeli 1. Najboljše rezultate doseže metoda SuperSimpleNet z 72% povprečno natančnostjo (AP). Na drugi strani se je med slabšimi obnesla metoda SegDecNet s povprečno natančnostjo 39%. Ocenjujemo, da je slabši rezultat deloma posledica pomanjkanja predučena na splošnih slikah, saj skoraj vse ostale metode s pred-naučeni utežmi dosežejo boljše rezultate (z izjemo DINOv2). Obe metodi za učenje uporabljata tudi segmentacijsko masko. Pri obeh smo preizkusili tudi brez učenja segmentacije, kar pa je rezultat poslabšalo. Segmentacija igra ključno vlogo predvsem pri SuperSimpleNet – brez nje model doseže več kot 13 odstotnih točk slabši rezultat ter tako zaostane za dvema sodobnima hrbtničnima metodama (ViT in Swin).

V primerjavi s sodobnimi hrbtničnimi metodami za klasifikacijo, le-te, z izjemo DINOv2, dosežejo boljši rezultat od SegDecNet, vendar so za vsaj 10 odstotnih točk slabše od SuperSimpleNet s segmentacijo. Med njimi je najboljši rezultat dosegla metoda ViT z 62% povprečne natančnosti, nato sledijo Swin v1 (58%), ConvNext (52%) ter DINOv2 z registri (45%). Med vsemi metodami je najslabši rezultat dosegla metoda DINOv2 (24%).

Rezultati tudi nakazujejo, da je problem detekcij poškodb na oblačilih relativno kompleksen. Najboljša preizkušena metoda doseže le okoli 70% povprečno natančnost z vsaj 30% napako priklica (27 izpuščenih od 98 poškodovanih oblek) ter 3% napačnih lažno pozitivnih (34 napačnih od 932). Glavni problem torej ostaja vizualni izgled poškodb. Ocenjujemo, da temu delno pripomore tudi nezadostna ločljivost kamere, saj obleka pogosto zaseda manjši del vidne površine kamere, medtem ko so na drugi strani poškodbe lahko pogosto tudi majhne, kar še bolj poudari problem ločljivosti. Nekaj primerov končnih klasifikacij in segmentacij za metodo SuperSimpleNet je

prikazanih tudi na Sliki 2.

## 6 Zaključek

V članku smo predstavili problematiko detektiranja površinskih napak na oblačilih v kontekstu reciklaže ter izvedli primerjalno analizo sodobnih metod globokega učenja na podmnožici slik *Second-Hand Fashion* [10] iz realnega primera reciklaže oblačil. Ovrednotili smo naši dve predhodno predstavljeni metodi za detekcijo splošnih napak na industrijskih izdelkih (SegDecNet [12, 1] ter SuperSimpleNet [11]), ter štiri sodobne tipe hrbtničnih arhitektur za splošno klasifikacijo slik. Eksperimenti so pokazali, da je problem kompleksnejši od detekcije napak v klasičnih industrijskih okoljih, predvsem zaradi raznolikosti oblačil, prisotnosti gub, slabe vidnosti poškodb ter omejene kakovosti slik. Med izbranimi pristopi je najboljšo uspešnost dosegla metoda SuperSimpleNet, ki združuje nadzorovano in nenadzorovano učenje ter omogoča hkratno segmentacijo in klasifikacijo.

Rezultati kažejo na potencial uporabe globokega učenja za avtomatsko detekcijo poškodovanih oblačil, a hkrati izpostavljajo tudi izzive, ki jih bo treba nasloviti za učinkovito uporabo v praksi. V prihodnje bi bilo smiselno raziskati tudi metode večnivojske segmentacije, kombinacijo z večsenzorskimi podatki ter vključevanje več modalnosti za bolj robustno klasifikacijo poškodovanih kosov.

## Literatura

- [1] Jakob Božič, Domen Tabernik, and Danijel Škočaj. Mixed supervision for surface-defect detection: from weakly to fully supervi-

**Zahvala:** To delo je bilo delno podprto s strani projekta ARIS J2-4457 (RTFM), raziskovalnega programa Računalniški vid (P2-0214) in z uporabo slovenskega superračunalniškega omrežja SLING.

- sed learning. *Computers in Industry*, 129, 2021.
- [2] Timothée Darcet, Maxime Oquab, Julien Mairal, and Piotr Bojanowski. Vision Transformers Need Registers. *12th International Conference on Learning Representations, ICLR 2024*, pages 1–21, 2024.
  - [3] Alexey Dosovitskiy, Lucas Beyer, Alexander Kolesnikov, Dirk Weissenborn, Xiaohua Zhai, Thomas Unterthiner, Mostafa Dehghani, Matthias Minderer, Georg Heigold, Sylvain Gelly, Jakob Uszkoreit, and Neil Houlsby. An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale. In *International Conference on Learning Representations*, 2021.
  - [4] Oquab et al. DINOv2: Learning Robust Visual Features without Supervision. *Transactions on Machine Learning Research*, 1:1–32, 24.
  - [5] Kaiming He, Xinlei Chen, Saining Xie, Yanghao Li, Piotr Dollar, and Ross Girshick. Masked Autoencoders Are Scalable Vision Learners. In *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, volume 2022-June, pages 15979–15988, 2022.
  - [6] Ze Liu, Han Hu, Yutong Lin, Zhuliang Yao, Zhenda Xie, Yixuan Wei, Jia Ning, Yue Cao, Zheng Zhang, Li Dong, Furu Wei, and Baining Guo. Swin Transformer V2: Scaling Up Capacity and Resolution. *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2022-June:11999–12009, 2022.
  - [7] Ze Liu, Yutong Lin, Yue Cao, Han Hu, Yixuan Wei, Zheng Zhang, Stephen Lin, and Baining Guo. Swin Transformer: Hierarchical Vision Transformer using Shifted Windows. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 9992–10002, 2021.
  - [8] Zhikang Liu, Yiming Zhou, Yuansheng Xu, and Zilei Wang. SimpleNet: A Simple Network for Image Anomaly Detection and Localization. *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2023-June:20402–20411, 2023.
  - [9] Zhuang Liu, Hanzi Mao, Chao-Yuan Wu, Christoph Feichtenhofer, Trevor Darrell, and Saining Xie. A ConvNet for the 2020s. In *Computer Vision and Pattern Recognition*, pages 11976–11986, 2020.
  - [10] Farrukh Nauman. Clothing Dataset for Second-Hand Fashion, 9 2024.
  - [11] Blaž Rolih, Matic Fučka, and Danijel Skočaj. SuperSimpleNet: Unifying Unsupervised and Supervised Learning for Fast and Reliable Surface Defect Detection. In *International Conference on Image Processing*, pages 47–65, 2025.
  - [12] Domen Tabernik, Samo Šela, Jure Skvarč, and Danijel Skočaj. Segmentation-Based Deep-Learning Approach for Surface-Defect Detection. *Journal of Intelligent Manufacturing*, 2020.
  - [13] Huosheng Xie and Zesen Wu. A robust fabric defect detection method based on improved refinedet. *Sensors (Switzerland)*, 20(15):1–24, 2020.
  - [14] Peizhe Zhao. SPFFNet: Strip Perception and Feature Fusion Spatial Pyramid Pooling for Fabric Defect Detection. 2025.
  - [15] Peng Zheng, Dehong Gao, Deng-Ping Fan, Li Liu, Jorma Laaksonen, Wanli Ouyang, and Nicu Sebe. Bilateral Reference for High-Resolution Dichotomous Image Segmentation. *CAAI Artificial Intelligence Research*, page 9150038, 2024.