

POST-HOC PAAIŠKINAMAS VAIZDŲ SEGMENTAVIMAS: TAIKYMAI INTERPRETUOJAMUMUI IR PRIEŠIŠKŲ ATAKŲ SCENARIJAMS

Parengė: Rokas Gipiškis

Vadovė: Prof. dr. Olga Kurasova

Doktorantūros pradžios ir pabaigos metai: 2020-2024 m. (4 metai)

DISERTACIJOS TEMA

Preliminari disertacijos tema studijų pradžioje:

LT: Konvoliuciniai neuroniniai tinklai galvos smegenų vaizdams segmentuoti

EN: Convolutional neural networks for the segmentation of human brain images

Disertacijos tema:

LT: Post-hoc paaiškinamas semantinis vaizdų segmentavimas: Taikymai interpretuojamumui ir priešišku atakų scenarijams

EN: Post-hoc explainable semantic image segmentation: Applications for interpretability and adversarial scenarios

EGZAMINAI

Studijų metai	Egzaminai	
	Planas	Įvykdyta
I (2020/2021)	2	2
II (2021/2022)	2	2
III (2022/2023)		
IV (2023/2024)		
Iš viso:	4	4

KONFERENCIJŲ IR PUBLIKACIJŲ PLANAS

Studijų metai	Dalyvavimas konferencijose				Publikacijos					
	Tarptautinėse		Nacionalinėse		Su citav. rodikliu			Be citav. rodiklio		
	Planas	Įvykdyta	Planas	Įvykdyta	Planas	Įvykdyta	Būklė	Planas	Įvykdyta	Būklė
I (2020/2021)										
II (2021/2022)			1	1				1	0	
III (2022/2023)	1	1			1	1	Publikuota	1	2	Publikuota
IV (2023/2024)	1	1			1	1	Publikuota			
Iš viso:	2	2	1	1	2	2		2	2	

STAŽUOTĖS

Stažuotės		
Vieta	Trukmė	Tyrimų grupė
Milano universitete, Italija	3 mėn.	Informatikos fakultetas, Industrial, Environmental and Biometric Informatics Laboratory
Neapolio universitete, Italija	5 mėn.	Matematikos fakultetas, M.O.D.A.L. tyrimų grupė
Padujos universitete, Italija	4 mėn.	Informacijos inžinerijos fakultetas, DEI tyrimų grupė
NSYSU universitete, Taivanas	5 mėn.	Informatikos ir inžinerijos fakultetas, Dirbtinio intelekto tyrimų grupė
Nacionalinis Taivano universitetas	5 mėn.	Informatikos ir informacijos inžinerijos fakultetas

DALYVAVIMAS KONFERENCIJOSE

Dalyvavimas tarptautinėse konferencijose	
	Aprašas
1.	R. Gipiškis, O. Kurasova “Occlusion-Based Approach for Interpretable Semantic Segmentation”. 18th Iberian Conference on Information Systems and Technologies (CISTI’2023), 2023 m. birželio 20-23 d., Portugalija, Aveiro.
2.	R. Gipiškis “XAI-driven Model Improvements in Interpretable Image Segmentation”. The 2nd World Conference on eXplainable Artificial Intelligence, 2024 m. liepos 17-18 d., Malta, Valeta.
Prisidėjimas ruošiant pranešimą tarptautinėje konferencijoje	
	Aprašas
1.	Gipiškis, Rokas, et al. "Ablation Studies in Activation Maps for Explainable Semantic Segmentation in Industry 4.0." IEEE EUROCON 2023-20th International Conference on Smart Technologies. IEEE, 2023 m. liepos 6-8 d., Italija, Turinas.

PUBLIKACIJOS

Publikacijos (tik su citavimo rodikliu)		
	Bibliografinis aprašas	Būklė
1.	Gipiškis Rokas, Chiaro Diletta, Preziosi Marco, Prezioso Edoardo, and Piccialli Francesco. "The impact of adversarial attacks on interpretable semantic segmentation in cyber–physical systems." <i>IEEE Systems Journal</i> 17.4 (2023): 5327-5334. DOI: 10.1109/JSYST.2023.3281079	Publikuota
2.	Gipiškis Rokas, Chun-Wei Tsai, and Olga Kurasova. "Explainable AI (XAI) in image segmentation in medicine, industry, and beyond: A survey." <i>ICT Express</i> (2024). DOI: 10.1016/j.ict.2024.09.008	Publikuota

DISERTACIJOS RENGIMO ETAPAI

3.	<p>Atskirų daktaro disertacijos dalių (tyrimo metodikos, rezultatų, ginamų teiginių, išvadų, ir kt.) parengimas:</p> <p>3.1. Tikslų, uždavinių, tyrimo metodikos, ginamųjų teiginių patikslinimas;</p> <p>3.2. Analitinės disertacijos dalies parengimas;</p> <p>3.3. Teorinės disertacijos dalies parengimas;</p> <p>3.4. Eksperimentinės disertacijos dalies parengimas;</p> <p>3.5. Bendrųjų išvadų formulavimas.</p>	2023 m. spalio mėn. – 2024 m. gegužės mėn.	<p>Atlikta išsami paaiškinamumo ir interpretuojamumo metodų semantiniame vaizdų segmentavime analizė. Pasiūlyta metodų taksonomija, atsižvelgiant į jų tipą, įverčio metriką bei taikymo sritį. Apžvelgtas paaiškinamo segmentavimo metodų taikymas medicinoje ir industrijoje. Identifikuoti pagrindiniai šios tyrimų srities iššūkiai.</p> <p>Parengtas ir publikuotas straipsnis, siūlantis interpretuojamo semantinio vaizdų segmentavimo taksonomiją. Ištirtos galimybės paaiškinamumo metodus pritaikyti instrumentiniams tikslams, t. y. siekiant pagerinti ne vien modelių interpretuojamumą, bet ir jų rezultatus.</p>
4.	Daktaro disertacijos parengimas ir svarstymas padalinyje	2024 m. birželio mėn.	Numatoma data: 2024 m. gruodžio mėn.
5	Daktaro disertacijos gynimas	2024 m. rugsėjo mėn.	Numatoma data: 2025 m. gegužės mėn.

TYRIMO OBJEKTAS IR TIKSLAS

Tyrimo objektas – giliojo mokymosi modelių **paaiškinamumo metodai semantinio vaizdų segmentavimo uždaviniui.**

Tikslas – sukurti **interpretuojamo semantinio vaizdų segmentavimo metodus**, pritaikomus konvoliuciniams neuroniniams tinklams.

UŽDAVINIAI

1. **Ištirti interpretuojamo** semantinio vaizdų **segmentavimo** ir klasifikavimo **metodus**, siekiant identifikuoti tinkamiausius sprendimus konvoliuciniams neuroniniams tinklams. Šio tyrimo pagrindu parengti pirmąjį šios srities apžvalginį straipsnį ir metodų taksonomiją.
2. **Integruoti** siūlomas **interpretuojamo** semantinio **segmentavimo modifikacijas**, leisiančias kokybiškai ir kiekybiškai **įvertinti** modelio išvesties rezultatai.
3. **Atlikti eksperimentinius tyrimus**, **tiriant** interpretuojamumo metodų **pritaikymą priešiškoje semantinio segmentavimo modelių atakose**, nukreiptose prieš paaiškinimus.

PUSMEČIO REZULTATAI

1. Parengtas ir **publikuotas straipsnis** su išsamia paaiškinamumo metodų semantiniame vaizdų segmentavime analize ir paaiškinamumo metodų taksonomija.

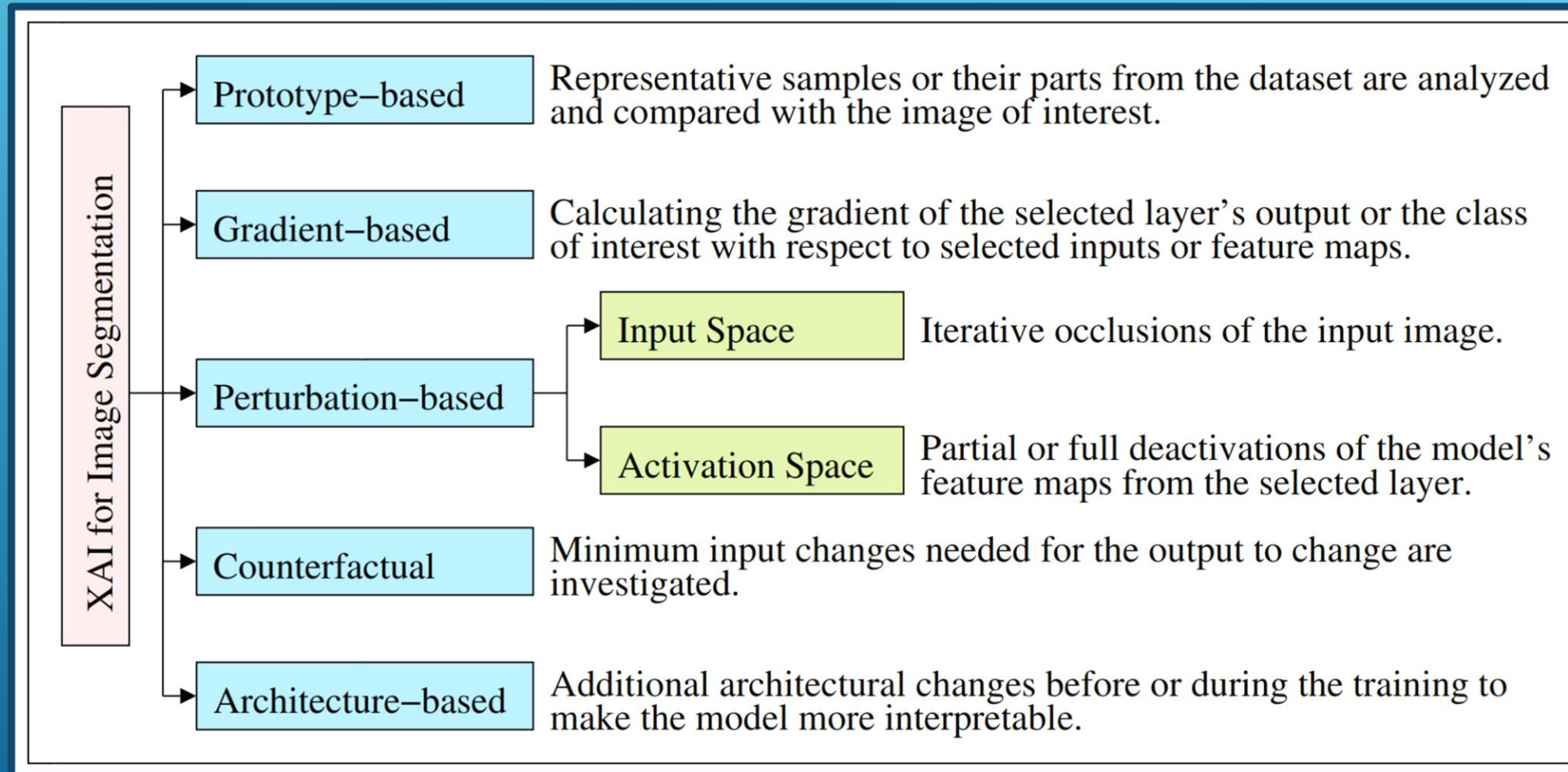
Gipiškis, Rokas, Chun-Wei Tsai, and Olga Kurasova. "Explainable AI (XAI) in image segmentation in medicine, industry, and beyond: A survey." *ICT Express* (2024).

2. **Parengtas ir pristatytas konferencijos pranešimas** apie semantinio segmentavimo interpretuojamumo metodų pritaikymo galimybes efektyvesnių giliojo mokymosi modelių kūrime.

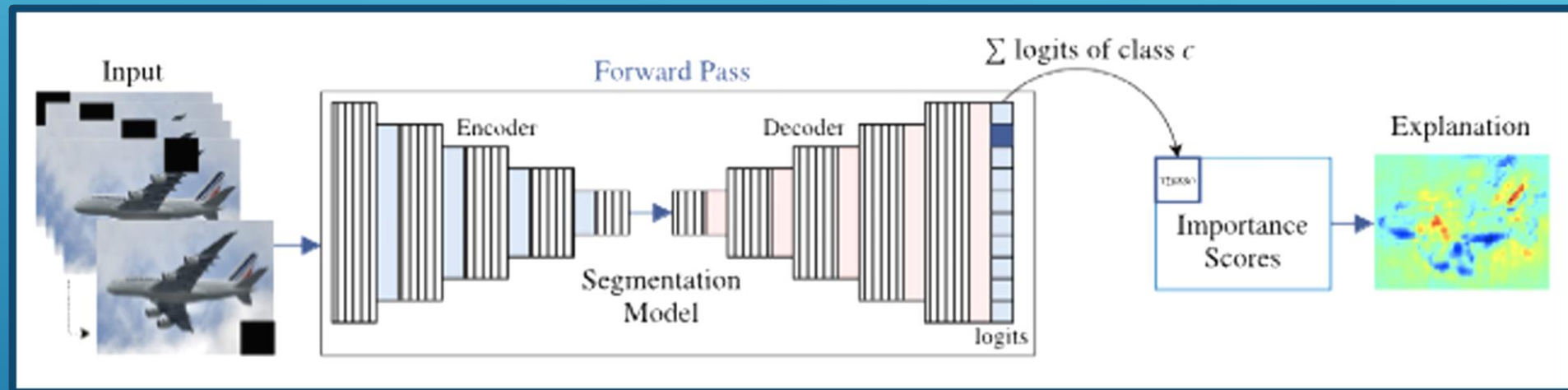
https://xaiworldconference.com/2024/xAI-2024:LB-D-DC/paper_47.pdf

(Numatoma publikuoti atviros prieigos CEUR-WS duombazėje)

PAAIŠKINAMO SEGMENTAVIMO TAKSONOMIJA



PERTURBACINIŲ PAAIŠKINIMŲ SCHEMA



[GTK24]

TAIKYMAI MEDICINOJE

Table 1

Explainable image segmentation in medicine.

Field	Imaging modality	Objects of interest	Datasets	Metric	Year	Ref.
G	IMG ^a	Colorectal polyps	EndoScene [97]	▷	2018	[14]
O	CT	Liver tumors	LiTS [98]	▶	2019	[99]
C	CMRI	Ventricular volumes	SUN09 [100], AC17 [101]	▷	2020	[102]
O	MRI	Brain tumors	TCGA	▷	2020	[103]
D	MRI/IMG	Skin lesions, multi-organ (incl. the fetal brain and the placenta)	ISIC2018 [104] and a private fetal MRI dataset	▷	2020	[105]
P	CT	Pancreatic region	Medical segmentation decathlon	▶	2021	[26]
C	CMRI	Ventricles, myocardium	Cardiac MRI dataset [101]	▷	2021	[106]
G	IMG	Polyps, med. instruments	Kvasir-SEG [78], Kvasir-Instrument [79]	▷	2021	[107]
O	MRI	Brain tumors	BraTS2018 [108]	▶	2021	[109]
FM	NIR	Iris	Private test dataset, post-mortem iris datasets, collected by [110]	▷	2022	[111]
V	CT/MRI/IMG	Skin lesions, abdomen multi-organ, brain tumors	HAM10000 [112], CHAOS 2019 [113], BraTS 2020 [108]	▷	2022	[114]
V	MRI	Brain tumors, human knees	BraTS 2017 [108], OAI ZIB [115]	▷	2022	[116]
O	MRI	Brain tumors	BraTS 2019 [108], BraTS 2021 [108]	▷	2022	[117]
N	MRA	Brain vessels	Private	▶	2022	[118]
H	IMG	Liver	Simulated dataset (Test Set 4) [119]	▷	2022	[120]
O	US/IMG	Breast tumors	Private LE/DES datasets, and BUSI [121]	▷	2023	[122]
G	CT/IMG	Colorectal polyps, lung cancer	EndoScene [97], LIDC-IDRI [123]	▶	2023	[124]
O	CT/MRI	Prostate cancer	3D pelvis dataset [125]	▷	2023	[126]
G	CT	Abdominal organs	Synapse multi-organ CT dataset [127]	▷	2023	[91]
O	BUS	Breast tumors	BUSI [121], BUSIS [128], HMSS [129]	▷	2023	[130]
O	CT/PET	Non-small cell lung cancer, whole-body	NSCLC, AutoPET [131]	▶	2023	[132]
D	IMG	Skin lesions	ISIC2018 [104]	▷	2023	[133]
D	IMG	Melanoma	ISIC 2018 [104], ISIC 2019 [112]	▷	2023	[134]
O	MRI/IMG	Prostate tumors, optic disc and cup	Prostate ^b and fundus ^c datasets	▷	2023	[135]
O	X-ray	Breast tumors	INbreast [136]	▶	2023	[137]
O	WSI	Head and neck tumors	Private	▷	2023	[138]
A	IR	Feet	ThermalFeet	▷	2023	[139]
V	CT/MRI/IMG	Brain tumors	BraTS 2018 [108], BraTS 2019 [108], BraTS 2020 [108], ISIC 2017	▷	2023	[140]
P	CT	Pancreas	Pancreas segmentation dataset [141]	▷	2023	[142]
Op	OCT	Retinal layers, glaucoma, diabetic macular edema	NR206, glaucoma dataset [143], DME dataset [144]	▷	2023	[145]
V	CT/MRI	Prostate, left ventricle, right ventricle, myocardium	NCI-ISBI 2013 [146], I2CVB [147], PROMISE12 [148], MSCMR [149], EMIDEC [150], ACDC [101], MMWHS [151], CASDC 2013 [152]	▷	2023	[153]
Op	OCT	Retinal layers, glaucoma, diabetic macular edema	Vis-105H, glaucoma dataset [143], DME dataset [144]	▷	2024	[154]
V	CMRI/CT	Left atrium, thoracic organs	Atrium dataset [141], SegTHOR [155]	▷	2024	[156]

A: Anesthesiology, C: Cardiology, D: Dermatology, FM: Forensic Medicine, G: Gastroenterology, N: Neurology,

O: Oncology, Op: Ophthalmology, P: Pancreatology, and V: Various.

▷: Qualitative XAI evaluation.

▶: Quantitative XAI evaluation.

^a IMG: general-purpose digital image formats, such as JPEG.

^b Prostate datasets: RUNMC [146], BMC [146], HCRUDB [147], UCL [148], BIDMC [148], and HK [148].

^c Fundus datasets: DRISHTI-GS [157], RIM-ONE-r3 [158], and REFUGE [159].

TAIKYMAI INDUSTRIJOJE

Table 2

Explainable image segmentation in industry.

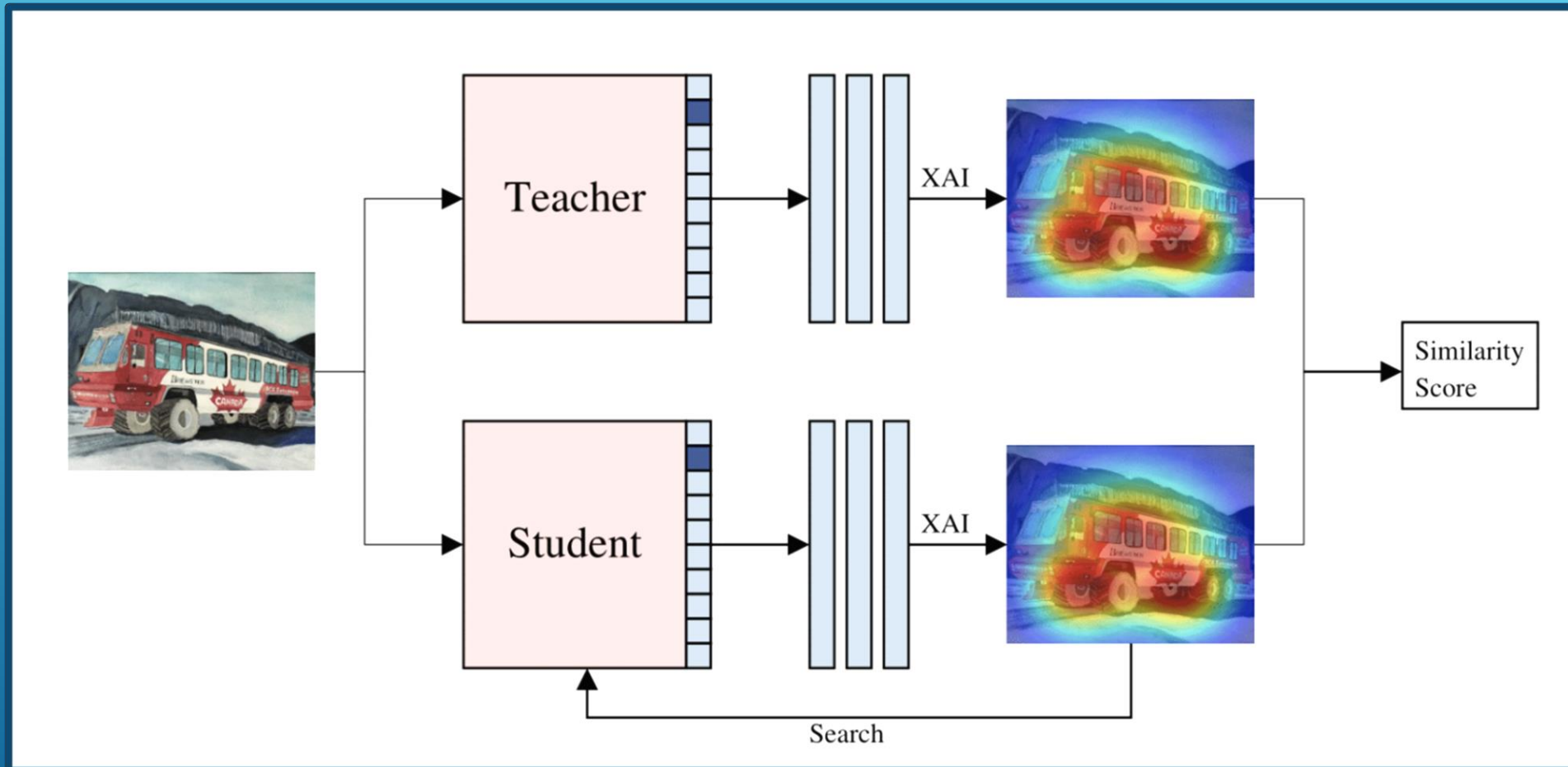
Category	Domain	Datasets	Metric	Year	Ref.
Remote sensing	Building detection	IAIL [162]	▷	2019	[163]
Scene understanding	Autonomous driving	SYNTHIA [164], A2D2 [165]	▷	2021	[166]
Scene understanding	Pedestrian environments	PASCAL VOC 2012 [65], ADE20K [167], Cityscapes [66]	NA ^a	2021	[168]
Scene understanding	Autonomous driving	KITTI [169]	▷	2022	[170]
Environmental monitoring	Flood detection	Worldfloods [71]	▷	2022	[62]
Scene understanding/Biometrics	Driving scenes/Face recognition	BDD100k [75], CelebAMask-HQ [171], CelebA [172]	▷	2022	[80]
Monitoring/Scene understanding	Drones/Food processing	ICG drone dataset, private dataset	▷	2023	[89]
Monitoring/General applications	Food processing	COCO [38], private dataset	▷	2023	[81]
Biometrics	Facial emotions	Face recognition dataset [173]	▷	2023	[174]
Monitoring	Cracks in infrastructure	CrackInfra [175]	▷	2023	[175]
General applications	Common objects	COCO [38]	▶	2023	[88]
Scene understanding/General applications	Street scenes/Common objects	Pascal VOC 2012 [65], Cityscapes [66]	▷	2023	[55]
Scene understanding	Driving scenes	BDD100k [75], BDD-OIA [76]	▷	2023	[74]
Scene understanding/General applications	Street scenes/Common objects	Cityscapes [66], Pascal VOC [65], COCO [38]	▶	2023	[176]
General applications	Common objects	COCO [38]	▷	2023	[91]
General applications	Common objects	Pascal VOC [65]	▶	2023	[124]
Scene understanding/Remote sensing	Street scenes/Building detection	Cityscapes [66], WHU [94]	▶	2023	[177]

▷: Qualitative XAI evaluation.

▶: Quantitative XAI evaluation.

^a The application focuses on introducing explainability to segmentation evaluation, rather than evaluating explainability techniques.

GALIMI PRITAIKYMAI NEURONINIŲ ARCHITEKTŪRŲ PAIEŠKOJE



[Gip24], remiantis [ZWJ23] pritaikymu klasifikavime

ŠALTINIAI

1. [GTK24] Gipiškis, Rokas, Chun-Wei Tsai, and Olga Kurasova. "Explainable AI (XAI) in image segmentation in medicine, industry, and beyond: A survey." *ICT Express* (2024).
2. [Gip24] Gipiškis, Rokas. "XAI-driven Model Improvements in Interpretable Image Segmentation", Late-breaking work, Demos and Doctoral Consortium, collocated with The 2nd World Conference on eXplainable Artificial Intelligence: July 17-19, 2024, Valletta, Malta (2024).
3. [ZWJ23] Zhang, Zhiyuan, Zhan Wang, and Inwhee Joe. "CAM-NAS: An Efficient and Interpretable Neural Architecture Search Model Based on Class Activation Mapping." *Applied Sciences* 13.17 (2023): 9686.