Artificial Intelligence for Better Prevention, Diagnosis, and Treatment of Health Conditions

Postępy w profilaktyce, diagnostyce i leczeniu schorzeń przy użyciu sztucznej inteligencji

Wyzwania i możliwości w rozwoju i wdrażaniu systemów AI
20-21 Listopada 2024 : AGH Kraków

Bogdan Matuszewski
Computer Vision and Machine Learning (CVML) Group
Institute for Engineering, Technology and Innovation (InETI)
School of Engineering and Computing
University of Central Lancashire

Acknowledgments

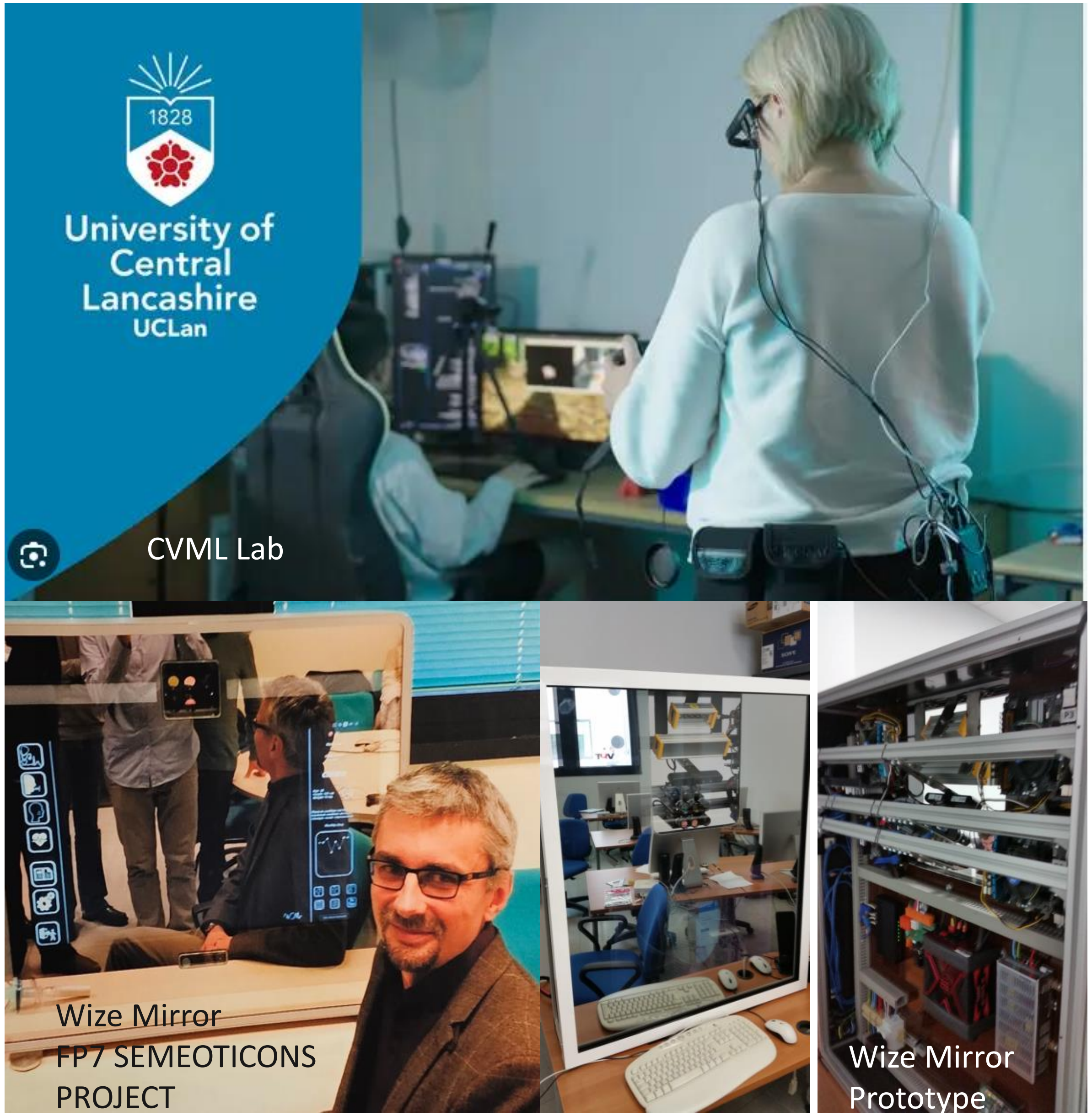
- Kerr Fitzgerald (UCLan; CVML PhD student)
- Dr Edward Sanderson (UCLan; School of Engineering and Computing); Former CVML PhD Student
- Dr Jianyu Zhou (UCLan; School of Engineering and Computing); Former CVML PhD Student
- Dr Katja Vogt (UCLan, School of Medicine and Dentistry)
- Dr Anastasia Topalidou (UCLan; School of Nursing and Midwifery)

- Professor James Turvill (Consultant Gastroenterologist at York & Scarborough Teaching Hospitals and Co-Chief Investigator for the national evaluations of the NHS England Colon Capsule Endoscopy (CCE) pilots)
- Professor Vishnu Chandrabalan (LTHTR, Director of the Lancashire and South Cumbria Secure Data Environment)
- Dr Charnley Natalie (LTHTR, Consultant in Clinical Oncology)
- Adnan A. Sheikh (ELHT, Consultant in Robotic and Minimally Invasive Colorectal and General Surgery)

Supported by the "AI demystified (AImyst)" project funded from QR Enhancing Research Culture Fund

1828
University of
Central Lancashire
UCLan

[UCLan Preston Campus Tour YouTube Link]



Institute for Engineering and Technology Innovation (InETI)

- Computer Vision and Machine Learning (CVML) Research Group
  - Group of 11 academics and 11 PhD students working in the areas of imaging, data analytics, AI (machine learning), Internet of Things (IoT), and mathematical modelling with applications in robotics, engineering, manufacturing, and health.
  - Managing Robotics and Vision Research Laboratory
  - Supporting MSc in Applied Data Science

- Historical note:
  - Applied Digital Signal and Image Processing (ADSIP) Research Centre established in 2001
  - Advanced Digital Manufacturing Technology (ADMT) Research Centre established in 2008
  - Computer Vision and Machine Learning (CVML) Research Group established in 2013 (https://www.uclan.ac.uk/research/institutes/ineti/cvml)

- Sample statistics (ADSIP+ ADMT+CVML, since 2013):
  - grant portfolio of over £7M (including EPSRC, STFC, MRC, EU, and industry funded),
  - collaboration with over 30 universities and companies nationally and internationally,
  - 17 PhD Completions,
  - 10 PDRA,
  - 11+ conferences/special sessions
    The most recent: Machine Learning in Endoscopy (EndoML) at MIUA’2024; 24-26 July 2024.

Selection of Active CVML Projects

- Information Monitoring System for personalised maternal health Treatment (IN-MOST)
- Analysing CT density heterogeneity in patients with metastatic renal cancer undergoing combination immunotherapy treatment.
- Facial recognition in acute illness – the facial study
- Artificial Intelligence systems for diagnostics of respiratory conditions (Intelligent Stethoscope)
- Development of early detection tools for sepsis diagnosis using artificial intelligence (Signal)

A representative sample of the outputs and outcomes resulting from our research

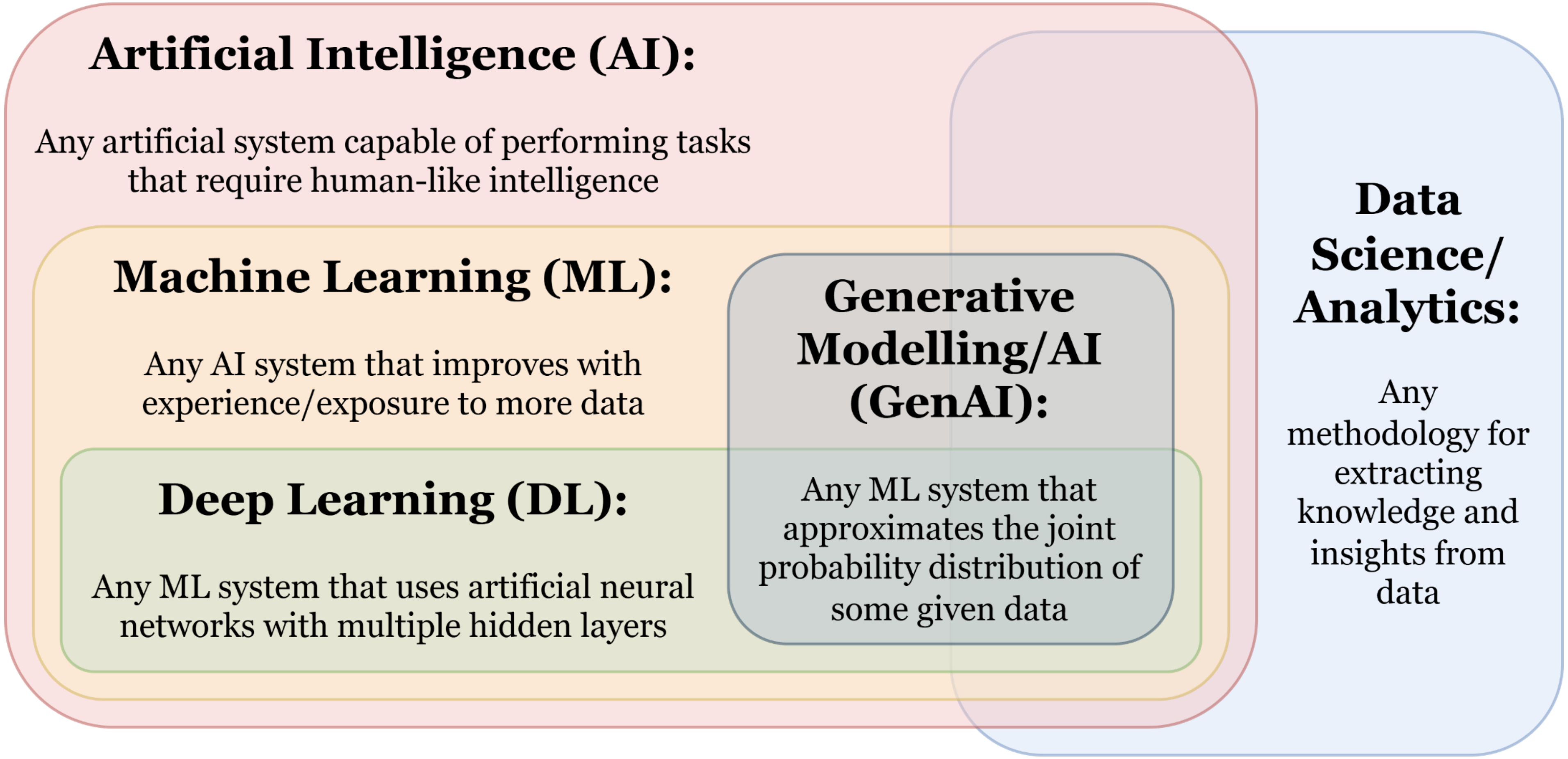
Suit of publicly available software:

- Deformable shape detection & modelling.
- AI for real-time 3D pose estimation.
- Polyp segmentation.
- Self-Supervised Pretraining for Vision Problems in Gastrointestinal Endoscopy for classification, detection, semantic segmentation, and monocular depth estimation.

Contribution to international challenges in bio-medical imaging

- Six award winning submissions, including:
- 1st place in "SimCol-to-3D: Simulated Colonoscopy data for 3D (scene) reconstruction" run as part of the MICCAI'2022 conference held in Singapore.
- 1st place in "Surgical Workflow Analysis in SensorOR" run as part of the MICCAI'2017 conference held in Quebec.
- 1st place in "Gastrointestinal Image ANAlysis (GIANA)" run as part of the MICCAI'2017 conference held in Quebec.

What is AI / Machine (Deep) Learning?



What is AI / Machine (Deep) Learning?

$$\{x\} \xrightarrow{F(x;a)=g_{\varphi}(f_{\theta}(x))?} y$$

x

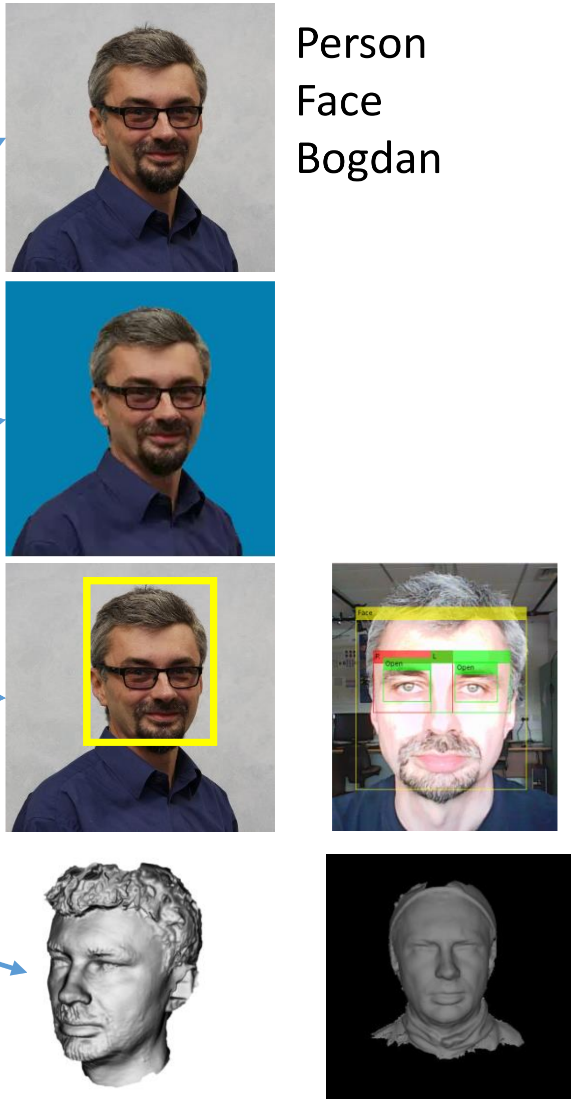
Recognition

Segmentation

Detection

3D Reconstruction

y



Supervised learning: $F(x,a)$

Self-Supervised representation learning: $f_{\theta}(x)$

Supervised learning recap

Supervised learning requires “labelled” data of the form (x, y)

We can train a model $g_\varphi(x) = \hat{y}$ to predict y from x by iterating over:

1. Evaluate the loss of a prediction $\hat{y}$ with respect to y
2. Update the parameters φ such as to reduce the loss

We call this “supervised” learning, as the provision of labels simulates a supervisory signal

A sample of current research

Deep Models for Endoscopy
Supported by the STFC CDN+ funded project:
"Machine Learning System for Decision Support and Computational Automation of Early Cancer Detection and Categorisation in Colonoscopy (**AIdDeCo**)"

[1] *Siegel, R.L., Miller, K.D., Fuchs, H.E., Jemal, A.: Cancer statistics, 2022. (2022)*
[2] *Salmo, E., Haboubi, N.: Adenoma and malignant colorectal polyp: pathological considerations and clinical applications. (2018)*
[3] *Kim, N.H., et al.: Miss rate of colorectal neoplastic polyps and risk factors for missed polyps in consecutive colonoscopies. (2017)*
[4] *Lee, J., et al.: Risk factors of missed colorectal lesions after colonoscopy. (2017)*
[5] *Corley, D.A., et al.: Adenoma detection rate and risk of colorectal cancer and death. (2014)*

Clinical Motivation

- Colorectal cancer (CRC) is one of the leading causes of cancer mortality worldwide [1]
- Commonly accepted that CRC develops from adenomatous polyps [2]
- Colonoscopy is the gold standard for colon screening and facilitates detection and treatment
- 17%-28% of colon polyps are missed during colonoscopy screening [3,4]
- Improvement of detection rates by 1% reduces risk of CRC by approximately 3% [5]

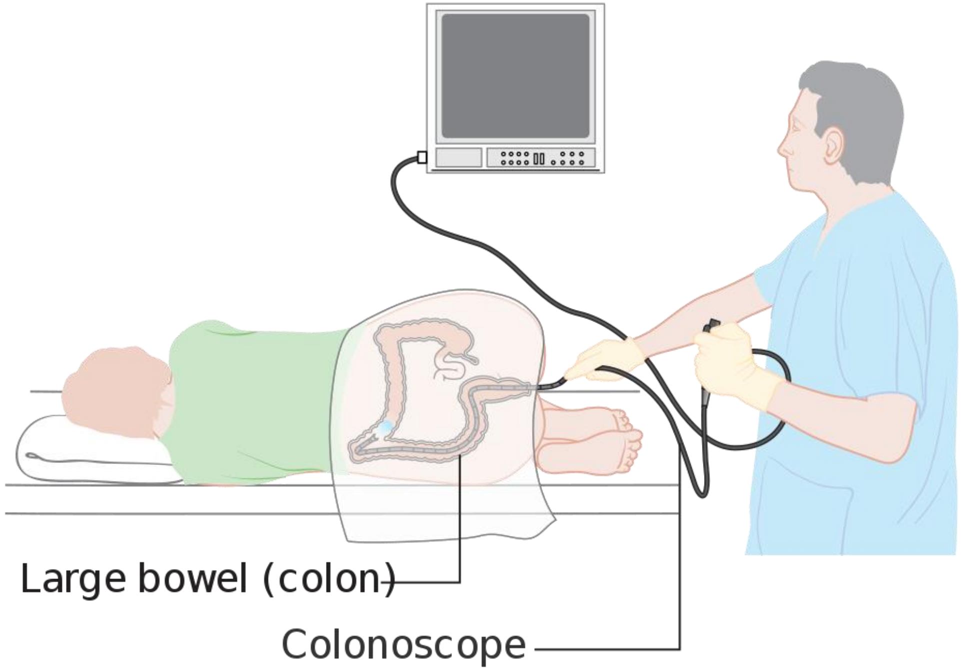


Diagram showing a colonoscopy CRUK 060 [Diagram]. Licensed under CC BY-SA 4.0.," Wikimedia Commons/Cancer Research UK, 2012. [Online]. https://commons.wikimedia.org/wiki/File:Diagram_showing_a_colonoscopy_CRUK_060.svg

AIdDeCo

Solutions

- Polyp Detection (CADe)
- Polyp Segmentation (CADe)
- Polyp Classification(CADx)
- Depth Recovery (CAQ)
- Navigation (CAQ)

- Colon Capsule Endoscopy with "Read Assist" and 3D colon reconstruction & navigation

CADe (Computer-aided polyp detection) - aims to decrease the rate of missed polyps during colonoscopy and ultimately increase the performance of the endoscopists

CADx (computer-aided diagnosis) – aims to perform a real-time polyp optical diagnosis, potentially being able to reduce the rate of unnecessary polypectomies of non-neoplastic lesions

CAQ (Computer-aided quality assessment) – aims to automatically assess performance during individual procedures

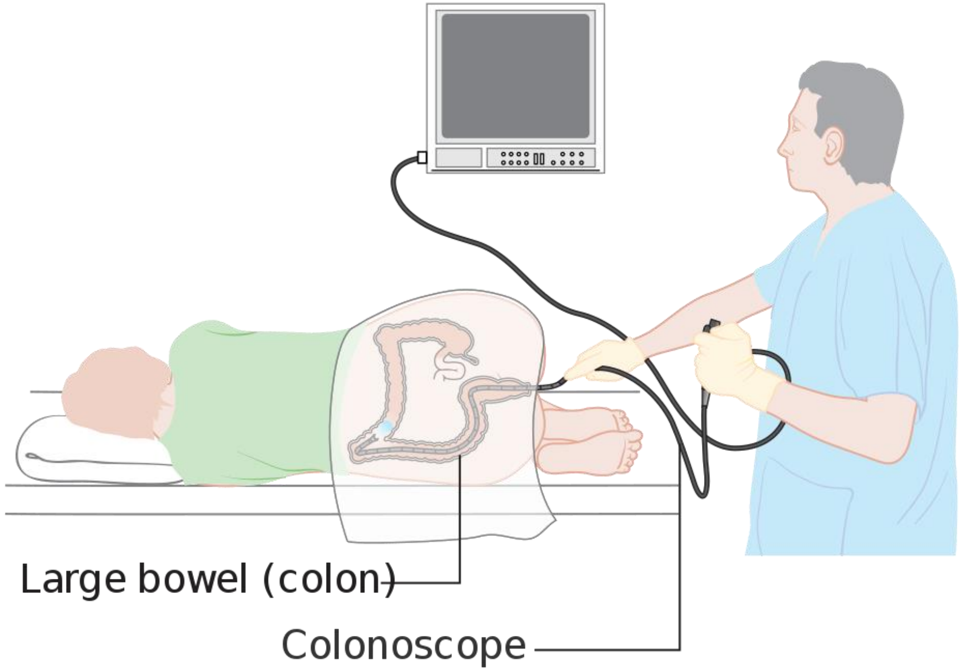


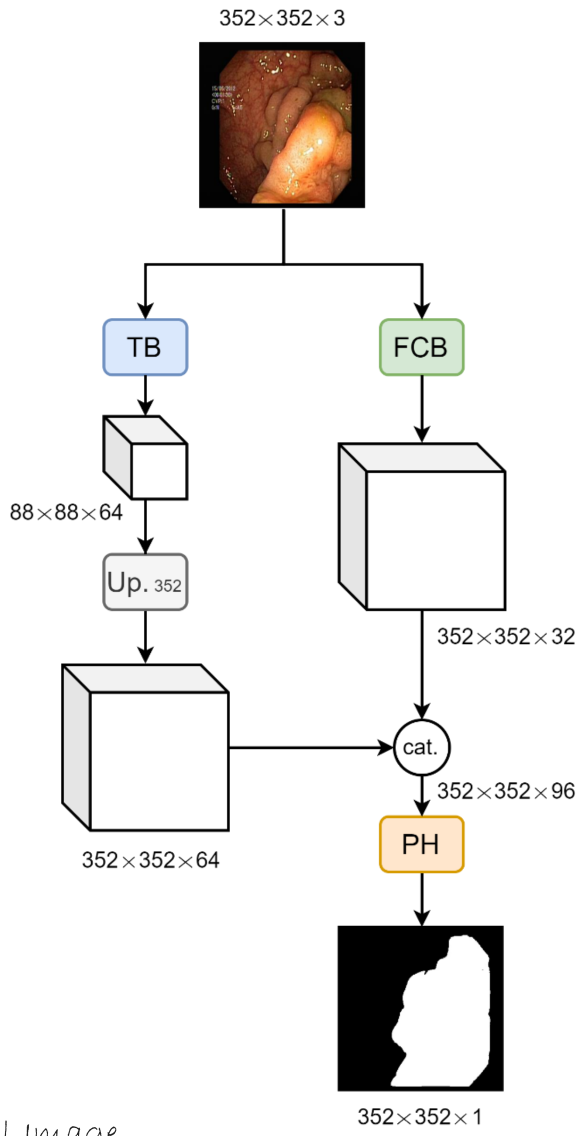
Diagram showing a colonoscopy CRUK 060 [Diagram]. Licensed under CC BY-SA 4.0.," Wikimedia Commons/Cancer Research UK, 2012. [Online]. https://commons.wikimedia.org/wiki/File:Diagram_showing_a_colonoscopy_CRUK_060.svg

AIdDeCo: Polyps Segmentation

FCBFormer

- Main components:
  - Transformer branch (TB)
  - Fully convolutional branch (FCB)
  - Up-sampling of TB output
  - Channel-wise concatenation of FCB output and up-sampled TB output
  - Prediction head (PH)

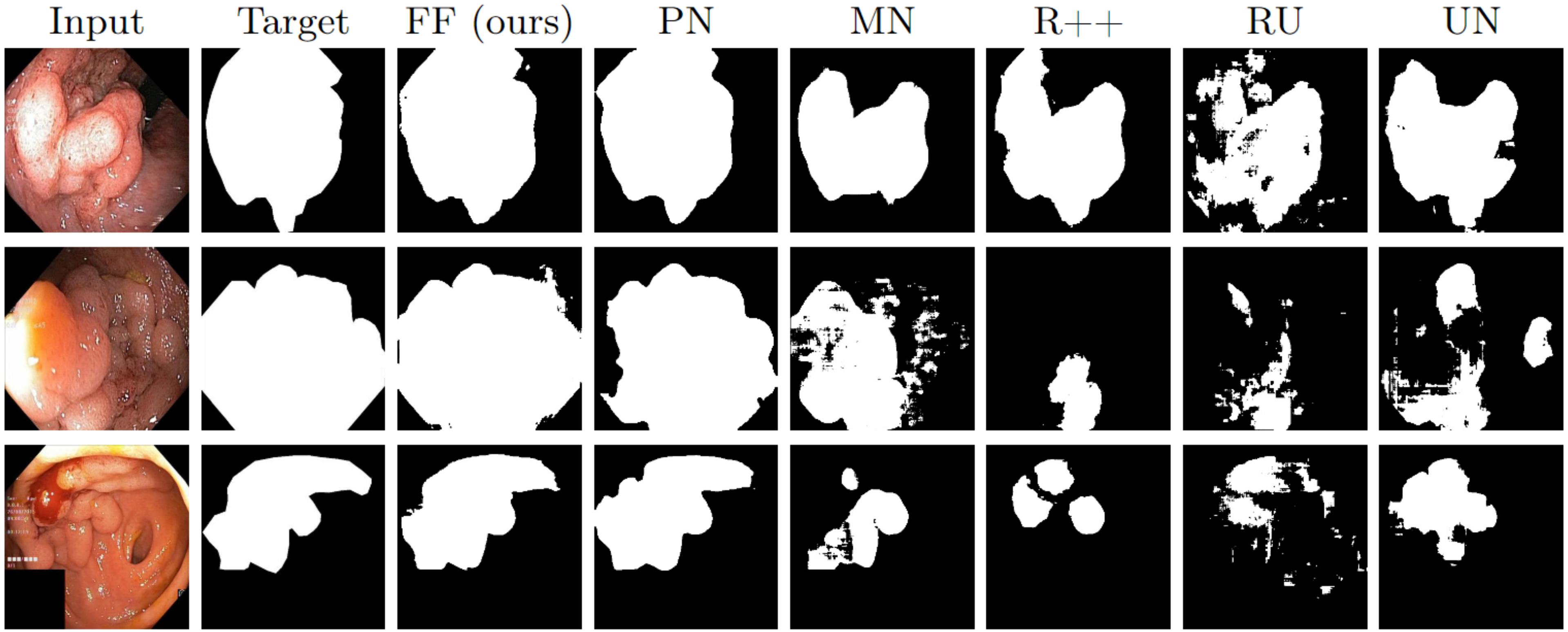
Code availability: https://github.com/CVML-UCLan/FCBFormer



Sanderson, E., Matuszewski, B.J. (2022). FCN-Transformer Feature Fusion for Polyp Segmentation. In: Medical Image Understanding and Analysis. MIUA 2022. Lecture Notes in Computer Science, vol 13413. Springer, Cham (CC BY 4.0)

AIdDeCo: Polyps Segmentation

Results: primary evaluation (challenging Kvasir-SEG examples)



Example inputs and targets from the Kvasir-SEG test set and the predictions for FCBFormer and the considered existing architectures. FF is FCBFormer, PN is PraNet, MN is MSRF-Net, R++ is ResUNet++, RU is ResUNet, and UN is U-Net. Each model used for this was the variant trained on the Kvasir-SEG training set.

Sanderson, E., Matuszewski, B.J. (2022). FCN-Transformer Feature Fusion for Polyp Segmentation. In: Medical Image Understanding and Analysis. MIUA 2022. Lecture Notes in Computer Science, vol 13413. Springer, Cham (CC BY 4.0)

AIdDeCo: Polyps Segmentation

Results from our experimentation

Training data	Kvasir-SEG				CVC-ClinicDB			
Test data	CVC-ClinicDB				Kvasir-SEG			
Metric	mDice	mIoU	mPrec.	mRec.	mDice	mIoU	mPrec.	mRec.
U-Net	0.5940	0.5081	0.6937	0.6184	0.5292	0.4036	0.4613	0.8481
ResUNet	0.3359	0.2425	0.5048	0.3307	0.3344	0.2222	0.2618	0.8164
ResUNet++	0.5638	0.4750	0.7175	0.5908	0.3077	0.2048	0.3340	0.4778
MSRF-Net	0.6238	0.5419	0.6621	0.7051	0.7296	0.6415	0.8162	0.7421
PraNet	0.7912	0.7119	0.8152	0.8316	0.7950	0.7073	0.7687	**0.9050**
FCBFormer (ours)	**0.8735**	**0.8038**	**0.8995**	**0.8876**	**0.8848**	**0.8214**	**0.9354**	0.8754

The same method gives different results.

There is a need to provide a measure of uncertainty for deep learning-based inference.

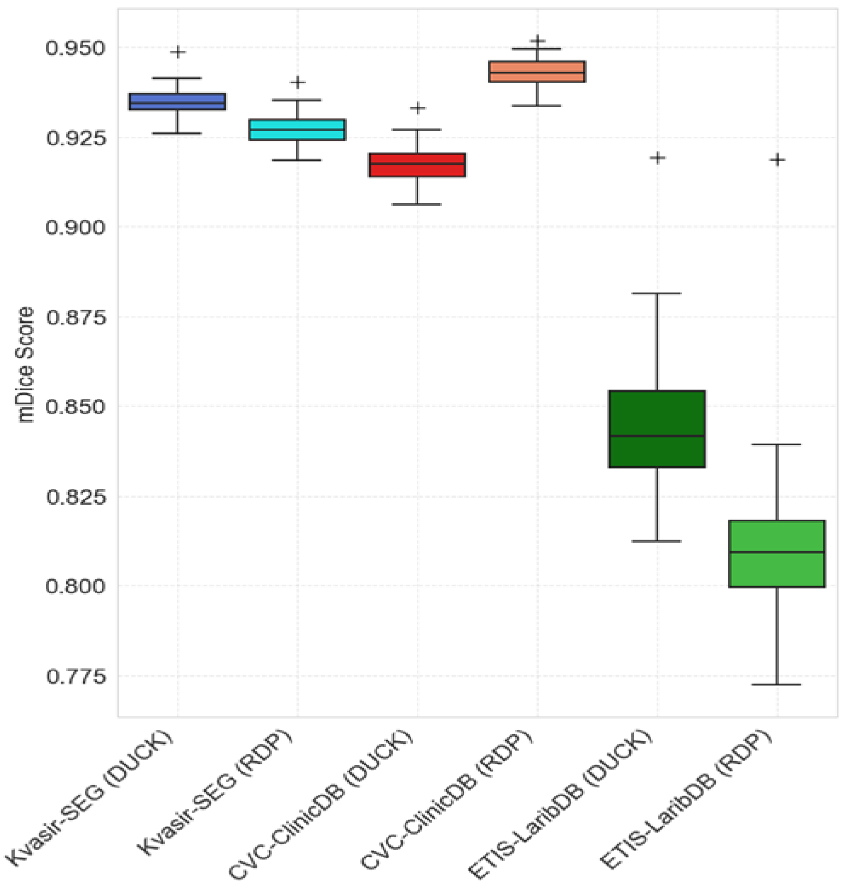
Results from literature (baselines) and our experimentation (FCBFormer)

Training data	Kvasir-SEG				CVC-ClinicDB			
Test data	CVC-ClinicDB				Kvasir-SEG			
Metric	mDice	mIoU	mPrec.	mRec.	mDice	mIoU	mPrec.	mRec.
U-Net	0.7172	0.6133	0.7986	0.7255	0.6222	0.4588	0.8133	0.5129
ResUNet++	0.5560	0.4542	0.6775	0.5795	0.5147	0.4082	0.7181	0.4860
MSRF-Net	0.7921	0.6498	0.7000	**0.9001**	0.7575	0.6337	0.8314	0.7197
PraNet	0.7225	0.6328	0.7888	0.7531	0.7293	0.6262	0.7623	0.8007
FCBFormer (ours)	**0.8735**	**0.8038**	**0.8995**	0.8876	**0.8848**	**0.8214**	**0.9354**	**0.8754**

Results:
generalisability tests

Sanderson, E., Matuszewski, B.J. (2022). FCN-Transformer Feature Fusion for Polyp Segmentation. In: Medical Image Understanding and Analysis. MIUA 2022. Lecture Notes in Computer Science, vol 13413. Springer, Cham (CC BY 4.0)

Note on Validation



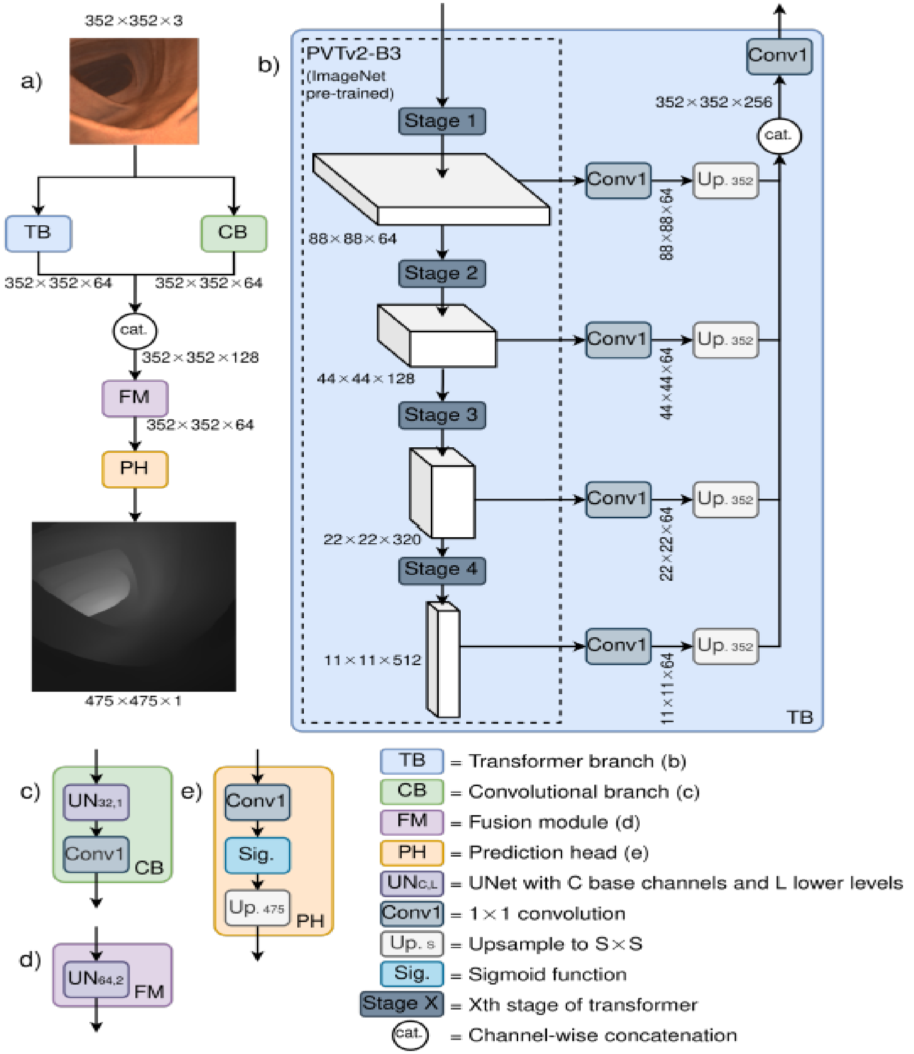
Boxplot visualisation of mDice score statistics from the 100 MC dropout runs conducted for each dataset using the DUCK-Net and random data partitions.

There is a need to provide a measure of uncertainty for deep learning-based inference.

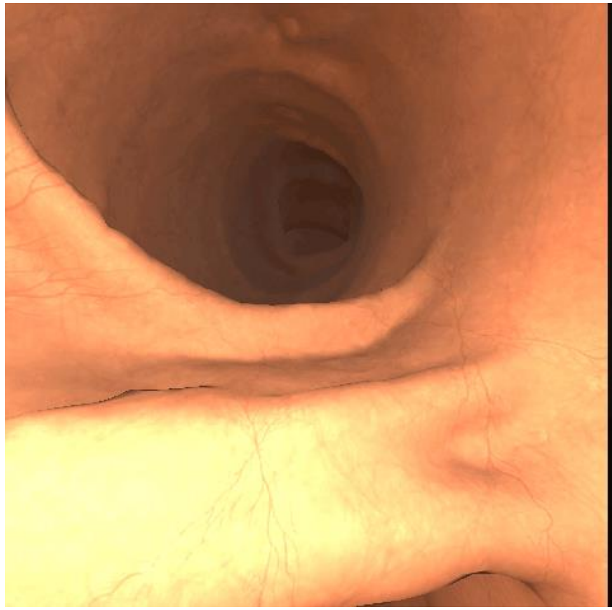
There are several ways to achieve this. Here is an example of using epistemic uncertainty estimation using the Monte Carlo dropout approach.

K. Fitzgerald, J. Bernal, A. Histace & B.J. Matuszewski, "Polyp Segmentation With the FCB-SwinV2 Transformer", IEEE Access, vol. 12, 2024

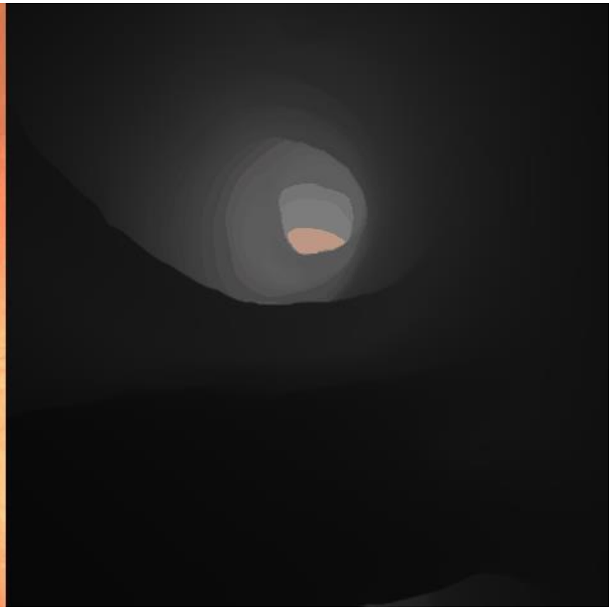
AIdDeCo: Depth Estimation



CVML submission to the SimCol3D Colonoscopy Challenge: Rau, A., et al., SimCol3D – 3D reconstruction during colonoscopy challenge, Medical Image Analysis Vol 96. August 2024. (https://clok.uclan.ac.uk/48117/ - Author-Accepted Manuscript available under CC-BY)



Input video - colonoscopy data



Depth estimation

CVML submission to the SimCol3D Colonoscopy Challenge: Rau, A., et al., SimCol3D – 3D reconstruction during colonoscopy challenge, Medical Image Analysis Vol 96. August 2024. (https://www.sciencedirect.com/science/article/pii/S1361841524001208) Author-Accepted Manuscript available under CC-BY)

Code availability :
https://github.com/ESandML/SimCol-Entry

Self-supervised learning (SSL)

"If artificial intelligence is a cake, self-supervised learning is the bulk of the cake" – Yann LeCun

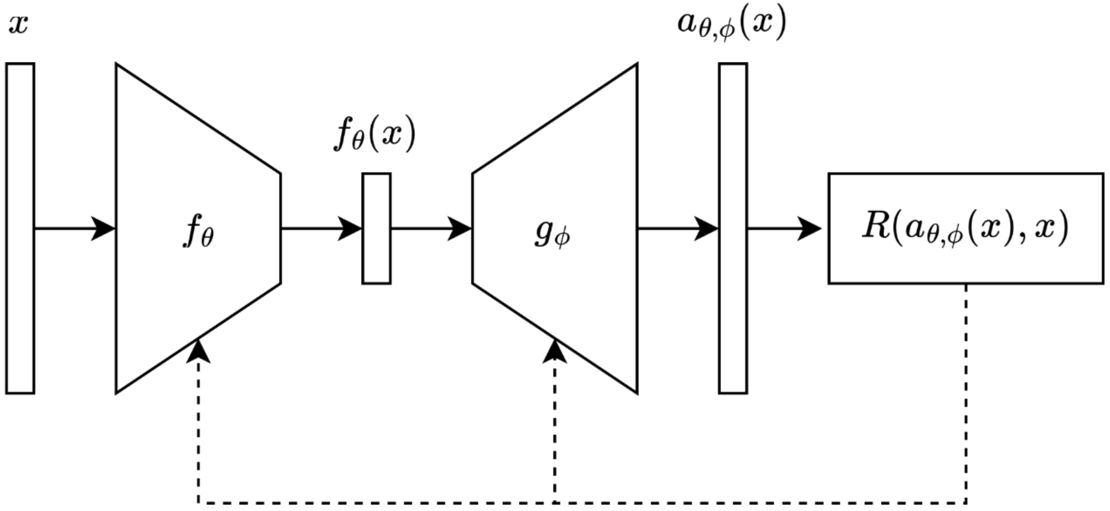
SSL is a way to train deep neural nets using only instances of the data x

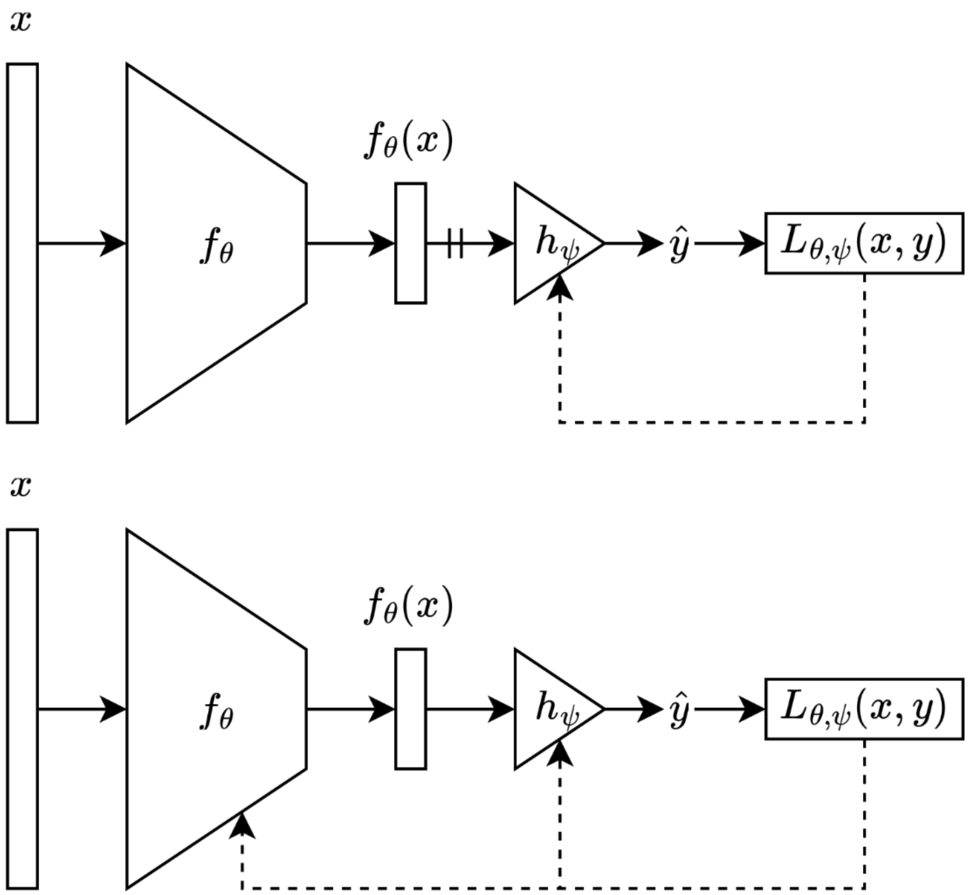
SSL algorithms train a network $f_\theta(x)$ on any of a wide range of (typically pretext) tasks:

- Autoencoding
- Contrastive learning
- Masked feature modelling
- Denoising
- Any other task which requires a model to "understand" the data

Although UL and SSL are operating in the same setup where the data labels are missing, the UL and SSL form a distinctive approaches.

Solving representation problem using self-supervised learning models trained on unlabelled data, with fine-tuned downstream application heads.





Step one:
Learning parameters of an encoder $f_\theta(x)$ in a self-supervised fashion; decoder g_ϕ , typically is not used after training si completed.

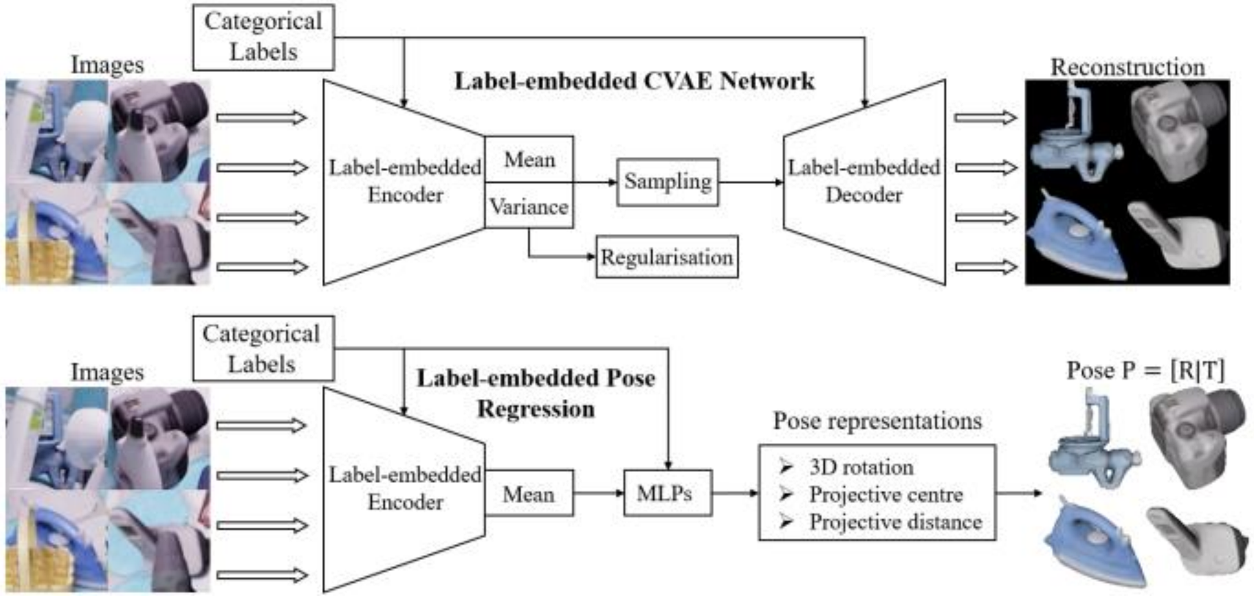
Step two:
Train a downstream application head h_ψ on a much smaller training set, possibly with fine tuning of the encoder.

A sample of current research

Deep Models for Real-Time Objects Characterisation

Supported by the EPSRC funded project:
"Self-Resilient Reconfigurable Assembly Systems with In-process Quality Improvement"

The scene understanding and automatic manipulation of objects present in that scene are of fundamental importance to robotics and automation. An image-based objects characterisation, e.g., shape, structure, weight, rigidity, pose..., is essential for many modern autonomous system.



J. Zhao, E. Sanderson and B. J. Matuszewski, "CVML-Pose: Convolutional VAE Based Multi-Level Network for Object 3D Pose Estimation," *IEEE Access*, vol. 11, pp. 13830-13845, 2023. (CC BY 4.0).

J. Zhao, W. Quan, and B. J. Matuszewski, CVAM-Pose: Conditional Variational Autoencoder for Multi-Object Monocular Pose Estimation, 35th British Machine Vision Conference 25th - 28th November 2024, Glasgow, UK (CC BY 4.0).

Images of objects are from Linemod and Linemod-Occluded Datasets.

Code availability:
https://github.com/JZhao12/CVML-Pose
https://github.com/JZhao12/CVAM-Pose

Example of a contrastive learning algorithm for computer vision:

[Advancing Self-Supervised and Semi-Supervised Learning with SimCLR – Google AI Blog (googleblog.com)]

The random distortion pipeline the authors found to work best was:

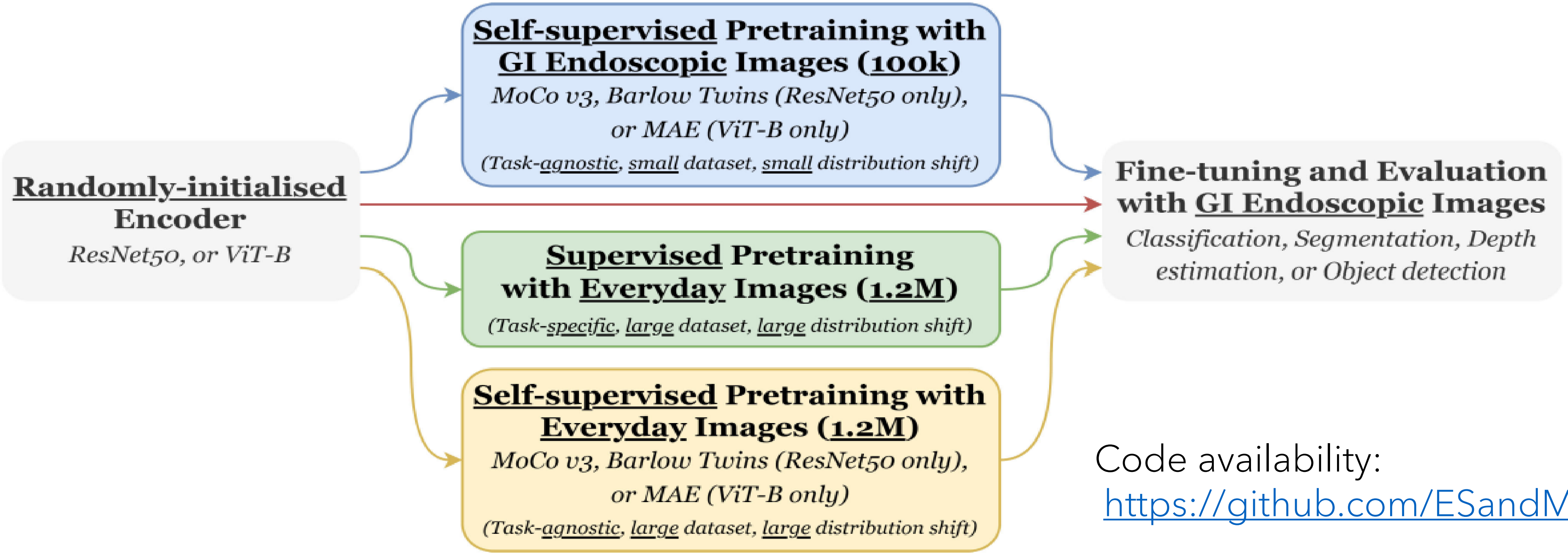
1. Random crop (with resize and random flip)
2. Random color distortion
3. Random Gaussian blur

Most subsequent contrastive learning algorithms for computer vision use this same pipeline

A sample of current research

[1] Chen, X., Xie, S., He, K.: An empirical study of training self-supervised vision transformers. (2021)
[2] Zbontar, J., Jing, L., Misra, I., LeCun, Y.: Barlow twins: Self-supervised learning via redundancy reduction
[3] He, K., Chen, X., Xie, S., Li, Y., Dollár, P., Girshick, R.: Masked autoencoders are scalable vision learners. (2022)
[4] Jha, D., et al.: Kvasir-seg: A segmented polyp dataset. (2020)
[5] Bernal, J., et al: WM-DOVA maps for accurate polyp highlighting in colonoscopy: Validation vs. saliency maps from physicians. (2015)

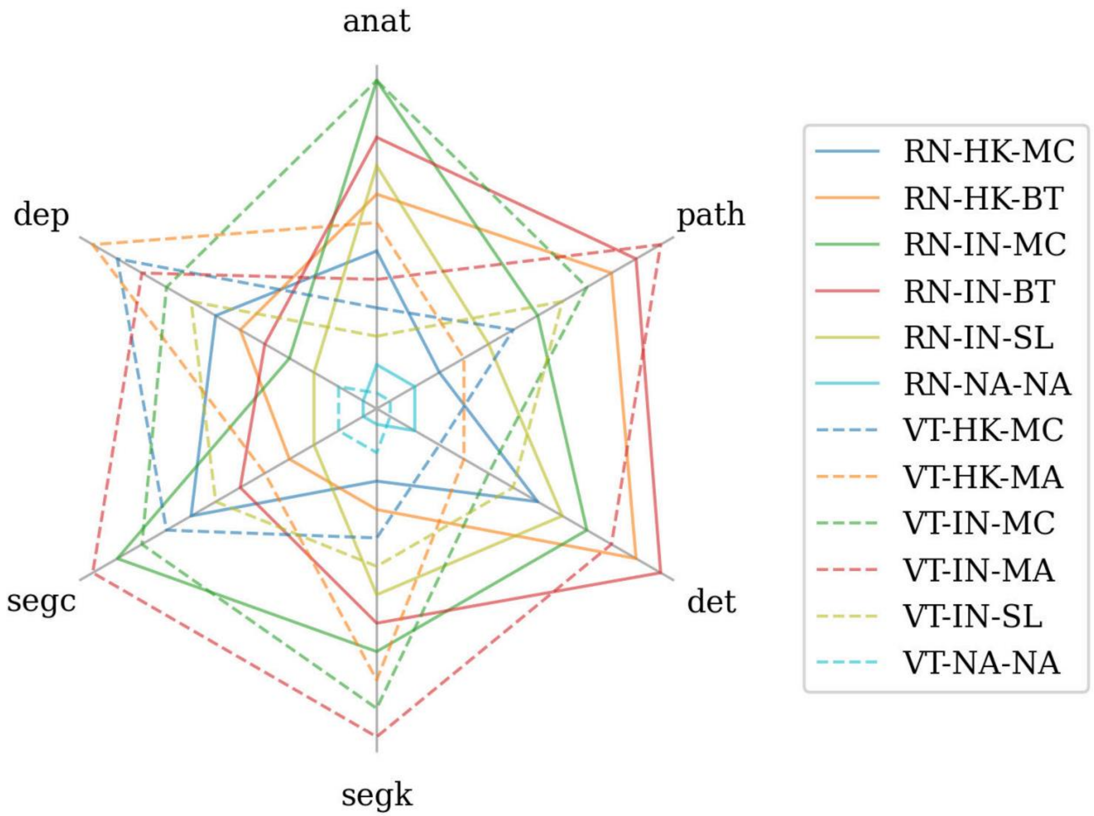
Self-Supervised Pretraining for Vision Problems in Gastrointestinal Endoscopy – is this useful?



Code availability:
https://github.com/ESandML/SSL4GIE

E. Sanderson and B. J. Matuszewski, "A Study on Self-Supervised Pretraining for Vision Problems in Gastrointestinal Endoscopy," in *IEEE Access*, vol. 12, pp. 46181-46201, 2024 (CC BY 4.0)

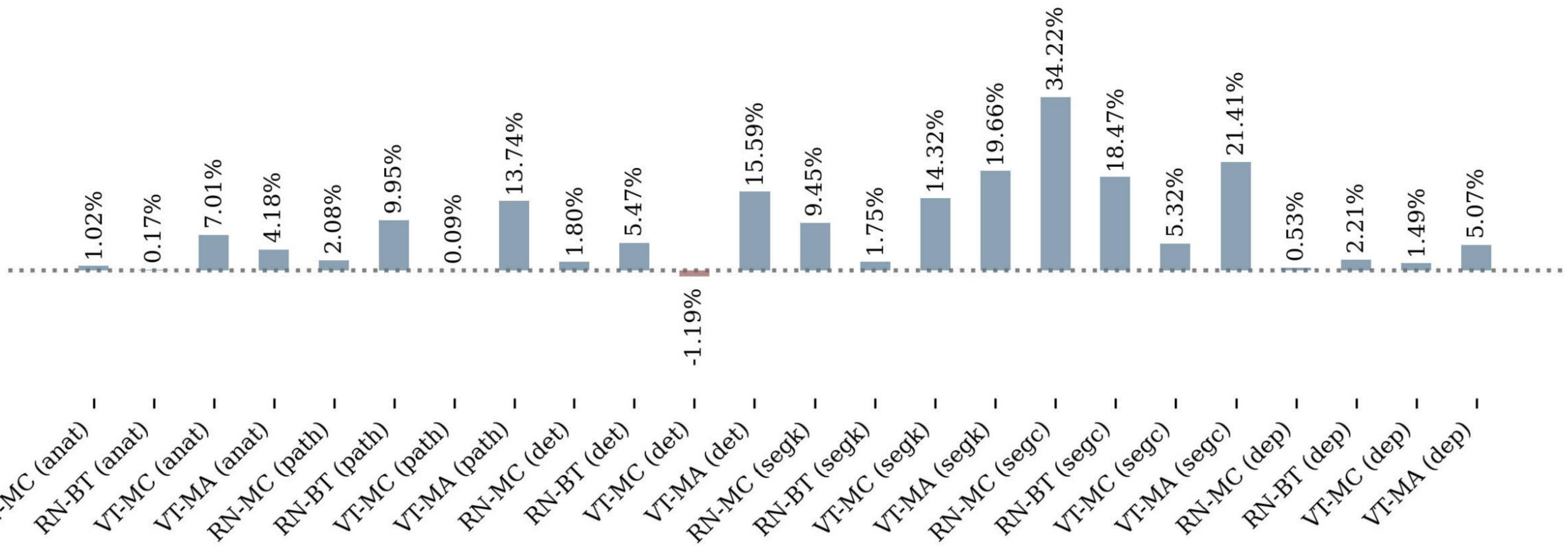
Self-Supervised Pretraining for Vision Problems in Gastrointestinal Endoscopy - Selected Results



Ranking of the performance of each model on each task, as measured by mF1 (anatomical landmark recognition and pathological finding characterisation), AP (polyp detection), mDice (polyp segmentation), and mRMSE (monocular depth estimation in colonoscopy), where a better rank is represented by a greater distance from the centre. For conciseness, we denote ResNet50s with RN, ViT-Bs with VT, Hyperkvasir-unlabelled with HK, ImageNet-1k with IN, MoCo v3 with MC, Barlow Twins with BT, MAE with MA, supervised pretraining with SL, and no pretraining with NA-NA. Additionally, we refer to anatomical landmark recognition as anat, pathological finding characterisation as path, polyp detection as det, polyp segmentation with Kvasir-SEG as segk, polyp segmentation with CVC-ClinicDB as segc, and monocular depth estimation in colonoscopy as dep.

E. Sanderson and B. J. Matuszewski, "A Study on Self-Supervised Pretraining for Vision Problems in Gastrointestinal Endoscopy," in *IEEE Access*, vol. 12, pp. 46181-46201, 2024 (CC BY 4.0)

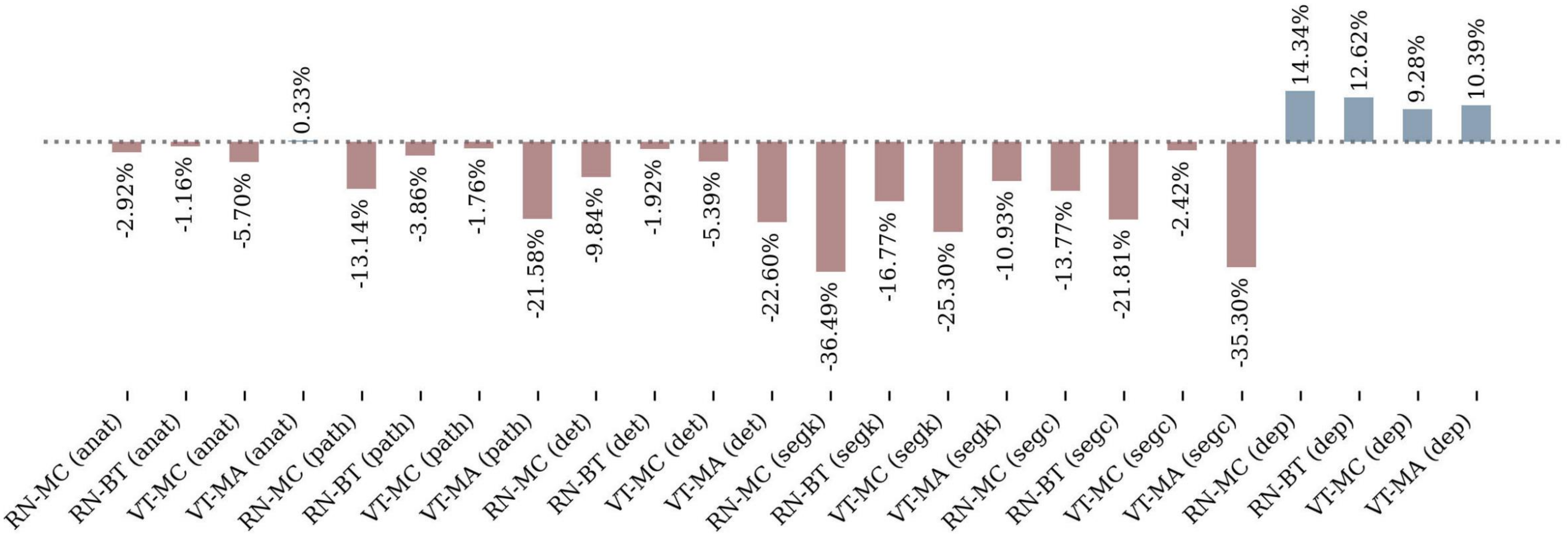
Self-Supervised Pretraining for Vision Problems in Gastrointestinal Endoscopy - Selected Results



Improvement of self-supervised pretraining vs. supervised pretraining for same architecture and pretraining data (ImageNet-1k). For conciseness, we denote ResNet50s with RN, ViT-Bs with VT, MoCo v3 with MC, Barlow Twins with BT, and MAE with MA. Additionally, we refer to anatomical landmark recognition as anat, pathological finding characterisation as path, polyp detection as det, polyp segmentation with Kvasir-SEG as segk, polyp segmentation with CVC-ClinicDB as segc, and monocular depth estimation in colonoscopy as dep.

E. Sanderson and B. J. Matuszewski, "A Study on Self-Supervised Pretraining for Vision Problems in Gastrointestinal Endoscopy," in *IEEE Access*, vol. 12, pp. 46181-46201, 2024 (CC BY 4.0)

Self-Supervised Pretraining for Vision Problems in Gastrointestinal Endoscopy - Selected Results



Improvement of pretraining with Hyperkvasir-unlabelled vs. pretraining with ImageNet-1k for same architecture and self-supervised pretraining algorithm. For conciseness, we denote ResNet50s with RN, ViT-Bs with VT, MoCo v3 with MC, Barlow Twins with BT, and MAE with MA. Additionally, we refer to anatomical landmark recognition as anat, pathological finding characterisation as path, polyp detection as det, polyp segmentation with Kvasir-SEG as segk, polyp segmentation with CVC-ClinicDB as segc, and monocular depth estimation in colonoscopy as dep.

E. Sanderson and B. J. Matuszewski, "A Study on Self-Supervised Pretraining for Vision Problems in Gastrointestinal Endoscopy," in *IEEE Access*, vol. 12, pp. 46181-46201, 2024 (CC BY 4.0)

Self-Supervised Pretraining for Vision Problems in Gastrointestinal Endoscopy - Selected Results

- Self-supervised pretraining generally produces more suitable backbones than supervised pretraining. This result is significant as it is still the convention to use backbones that have been pretrained on ImageNet-1k in a supervised manner. *This result contrasts with the results observed for tasks involving everyday images, where supervised pretraining typically leads to better performance.*

- Self-supervised pretraining with ImageNet-1k generally produces more suitable backbones than self-supervised pretraining with Hyperkvasir-unlabelled, with the notable exception of monocular depth estimation in colonoscopy where the similarity of the pretraining data to the downstream data appears to be more critical than the amount of pretraining data. *While this is a useful insight for the development of monocular depth estimation models for GIE, this finding may also be true for monocular depth estimation solutions in other domains.*

- That ResNet50 backbones are generally better for polyp detection, whereas ViT-B backbones are generally better for polyp segmentation and monocular depth estimation in colonoscopy, and both architectures perform similarly in anatomical landmark recognition and pathological finding characterisation.

Artificial Intelligence for Better Prevention, Diagnosis, and Treatment of Health Conditions

Postępy w profilaktyce, diagnostyce i leczeniu schorzeń przy użyciu sztucznej inteligencji

Wyzwania i możliwości w rozwoju i wdrażaniu systemów AI
20-21 Listopada 2024 : AGH Kraków

Bogdan Matuszewski
Computer Vision and Machine Learning (CVML) Group
Institute for Engineering, Technology and Innovation (InETI)
School of Engineering and Computing
University of Central Lancashire