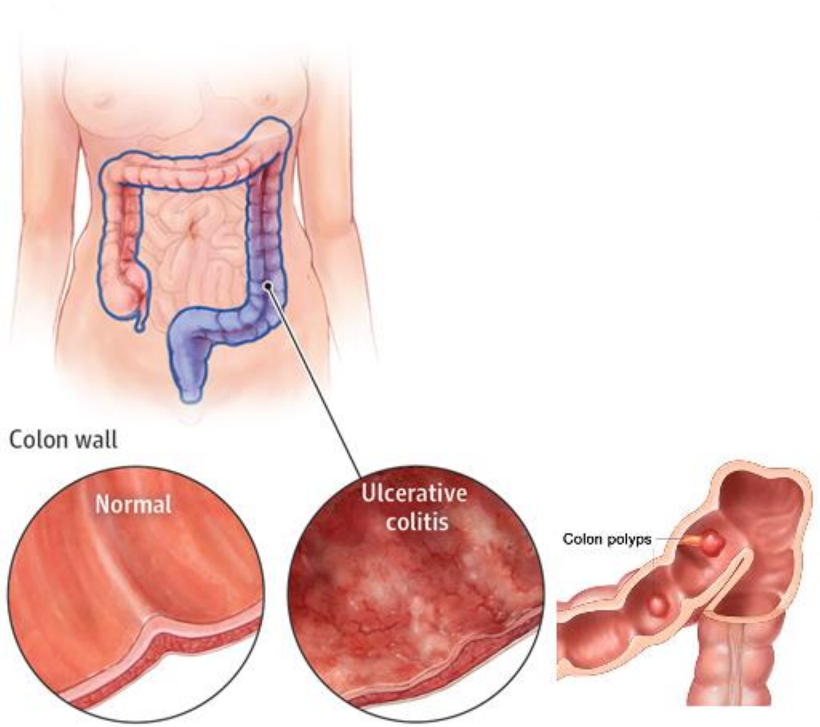
Automatic Pathology Detection

Ahmed Mohammed
Department of Computer Science
Norwegian University of Science and Technology
Sintef Digital

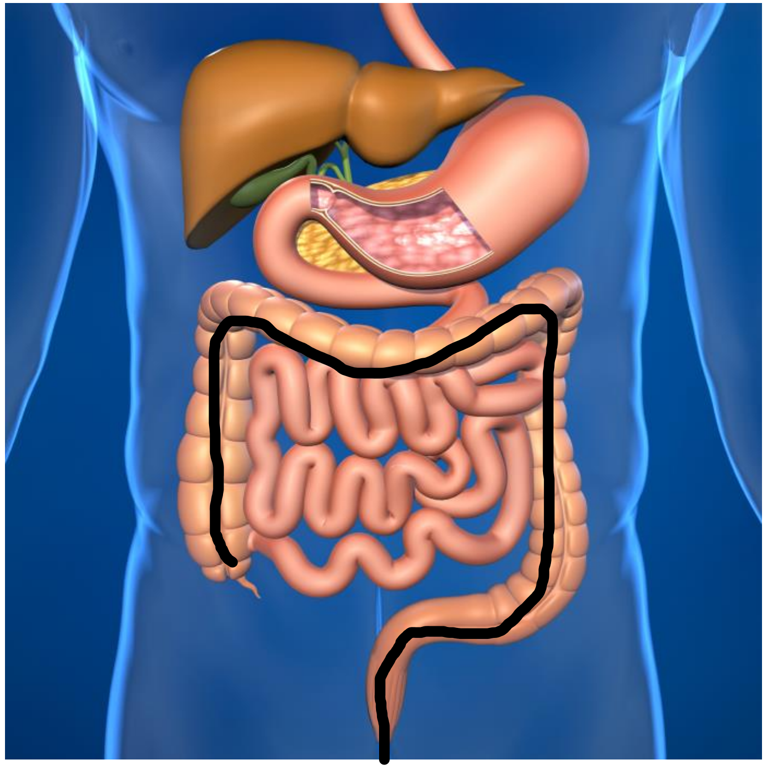
Motivation



Courtesy of: World Journal of Gastroenterology

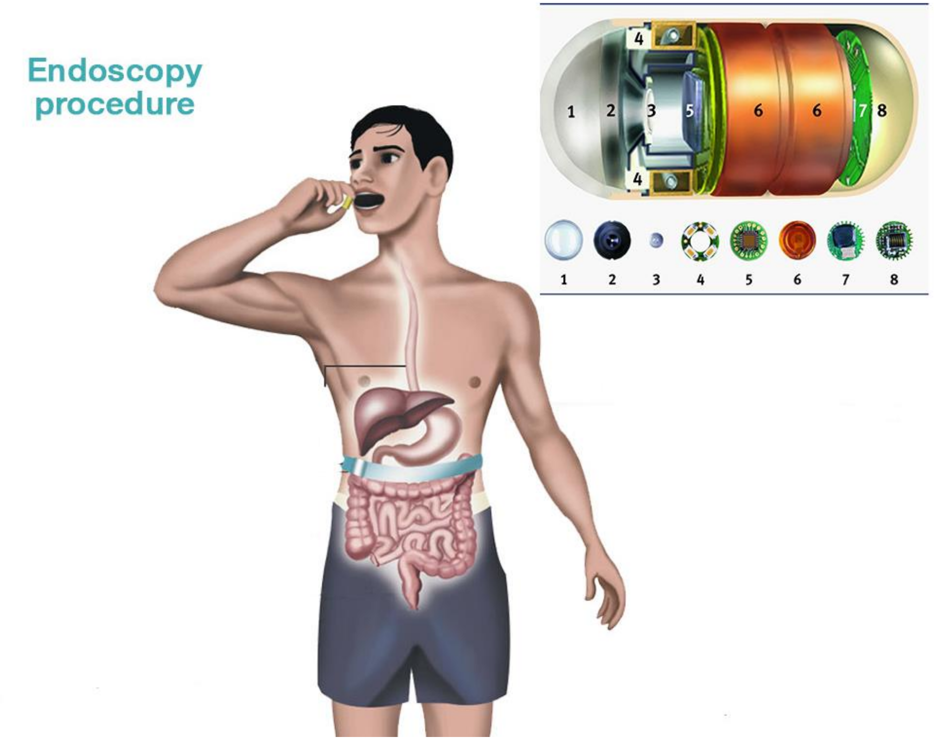
- Colorectal cancer (CRC) is a major cause of mortality and morbidity worldwide.
- Patients with long standing ulcerative colitis are at greater risk of developing colon or rectal cancer than the general population

Motivation

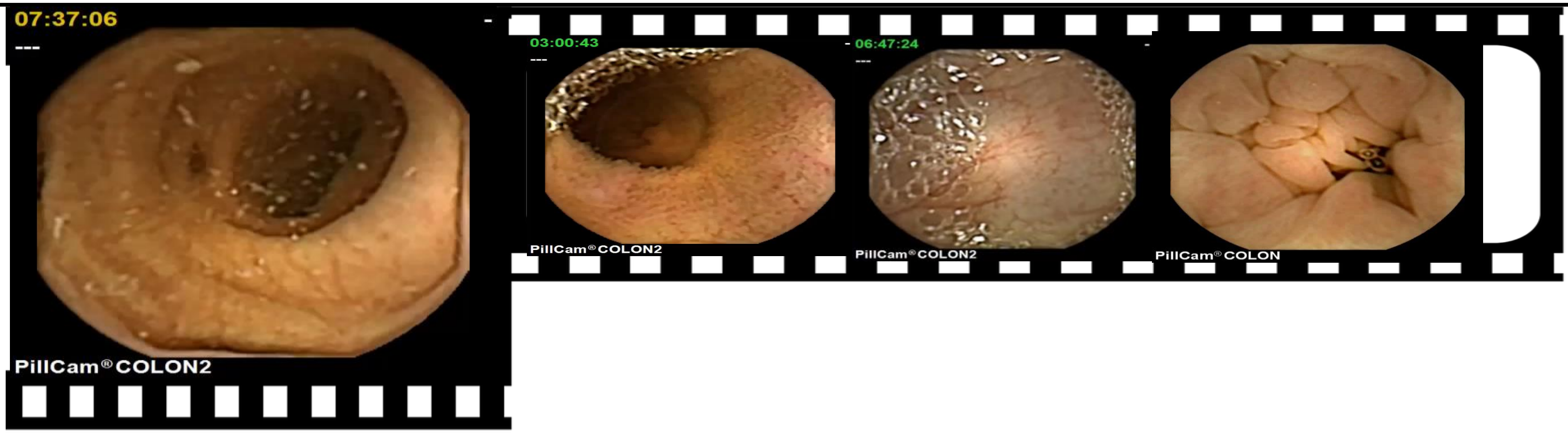


Capsule Video Endoscopy(CVE): Characteristics

- Size 31x11 mm
- Images produced (~50,000)
- requires 2 hours to review

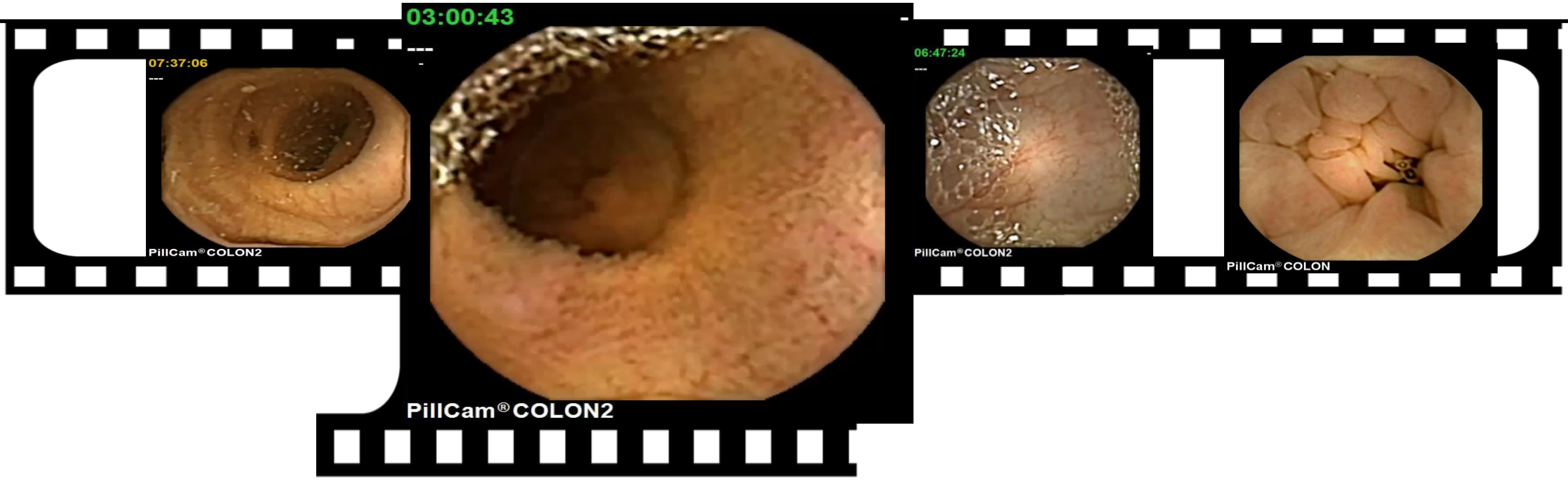


Capsule Video Endoscopy(CVE): Characteristics



IDEALLY
- Vessels should be visible
- Specular reflections should be removed
- Tissues should be visible
- Color shouldn't be distorted. (If the colors are changed, it requires retraining physician to view such images)

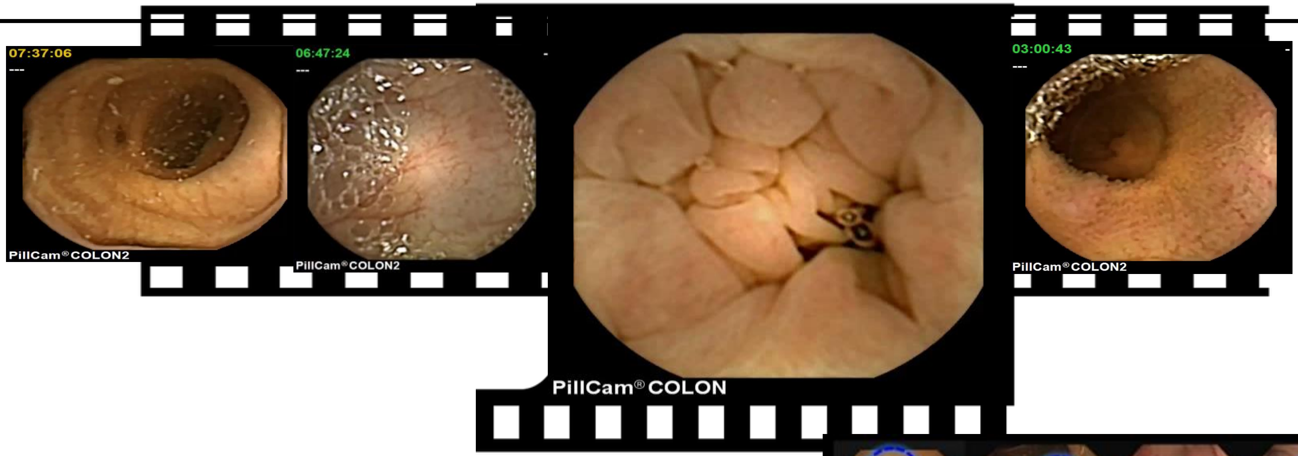
Capsule Video Endoscopy(CVE): Characteristics

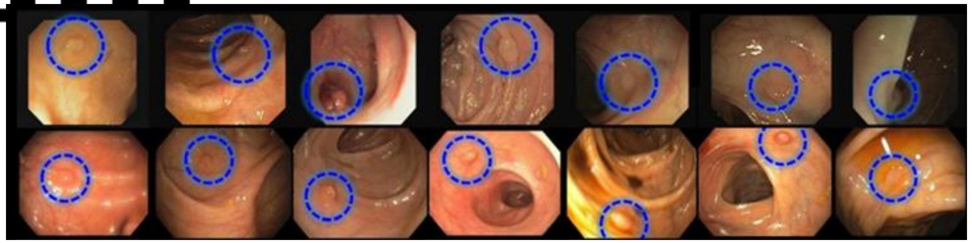


IDEALLY

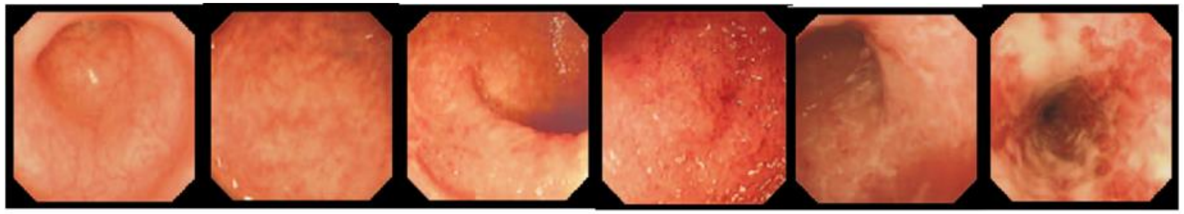
- The phi-effect should be obtained to create apparent living pictures with natural motion portrayal.
- Flickering should be avoided when displaying image sequences.
- Save battery

Capsule Video Endoscopy(CVE): Characteristics

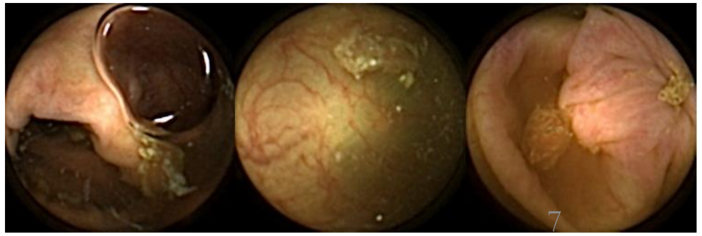




- Sessile
- Pedunculated
- Neoplastic vs Hyperplastic



Different stages of UC and their appearances



Appearance, orientation and shape variability,

Unstructured ⇒ Structured data

Challenge:

- Medical data annotation is *expensive and slow*
- Privacy issues and limited access
- Diverse pathologies and long video sequence

How to deal with lack of data

- Transfer Learning, Domain Adaptation
- Weakly supervised learning
- Self-supervision
- Semi- or Unsupervised learning

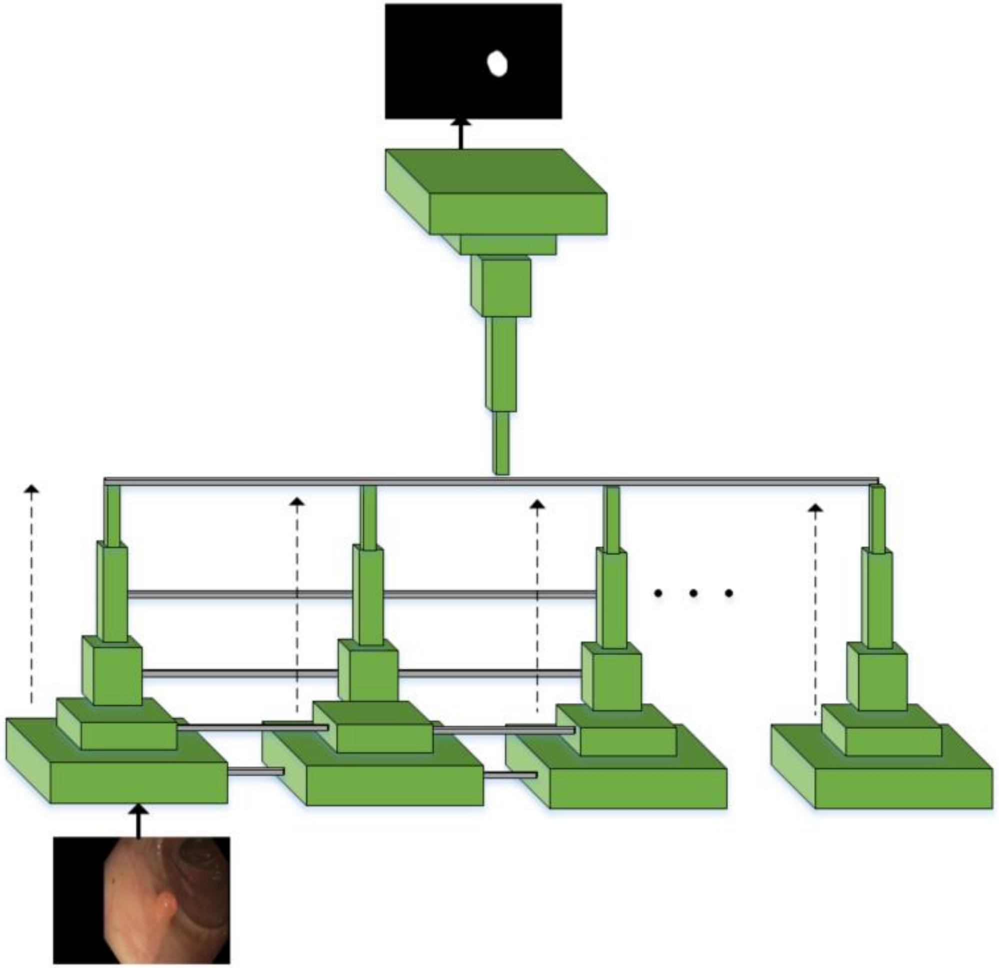
Approach

- Hand-crafted features approach:
  - Shape and appereance modeling: Intensity valley detection
  - Gives high FP and low TP
- Deep learning approach
  - Hybrid: CNN + RGB+ geometric features
  - CNN with RGB input, transfer learning and trainig from scratch
  - 3D CNN- Online and offline

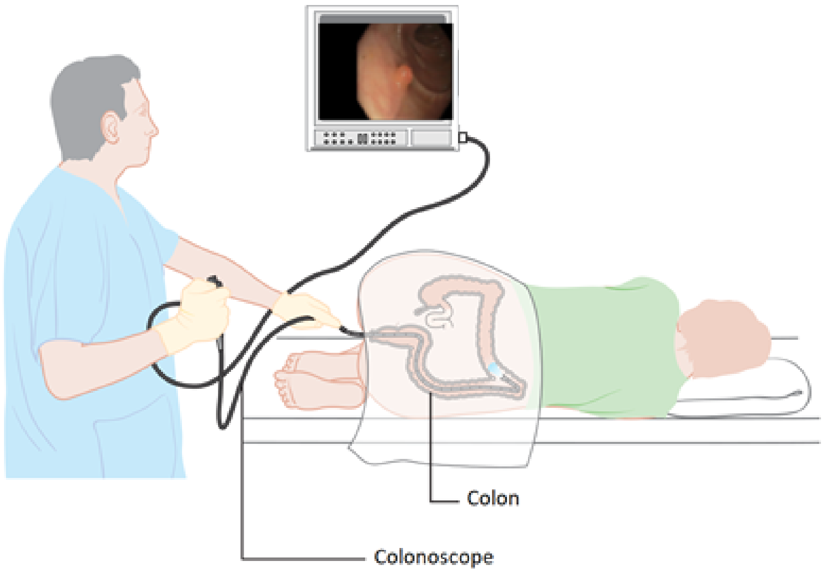
Proposed: Y-Net

- Exploit the advantage of pre-trained weights
- Combines pre-trained weights with random weights
- Address the performance loss due to domain-shift

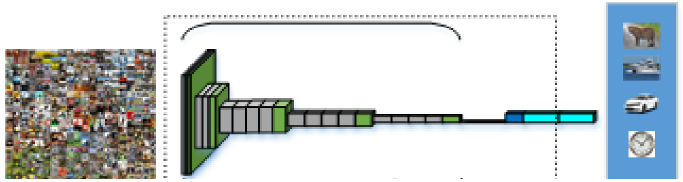
Proposed: Ensemble of encoders:

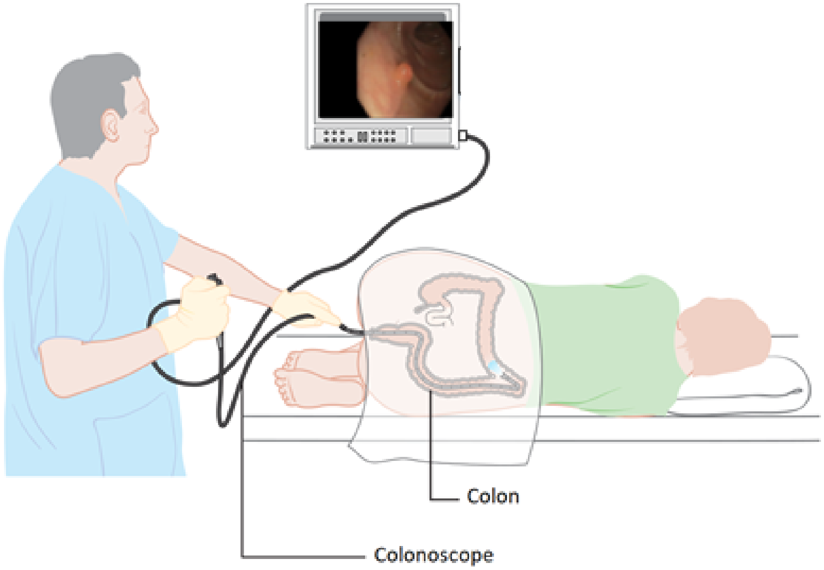


Proposed: Y-Net

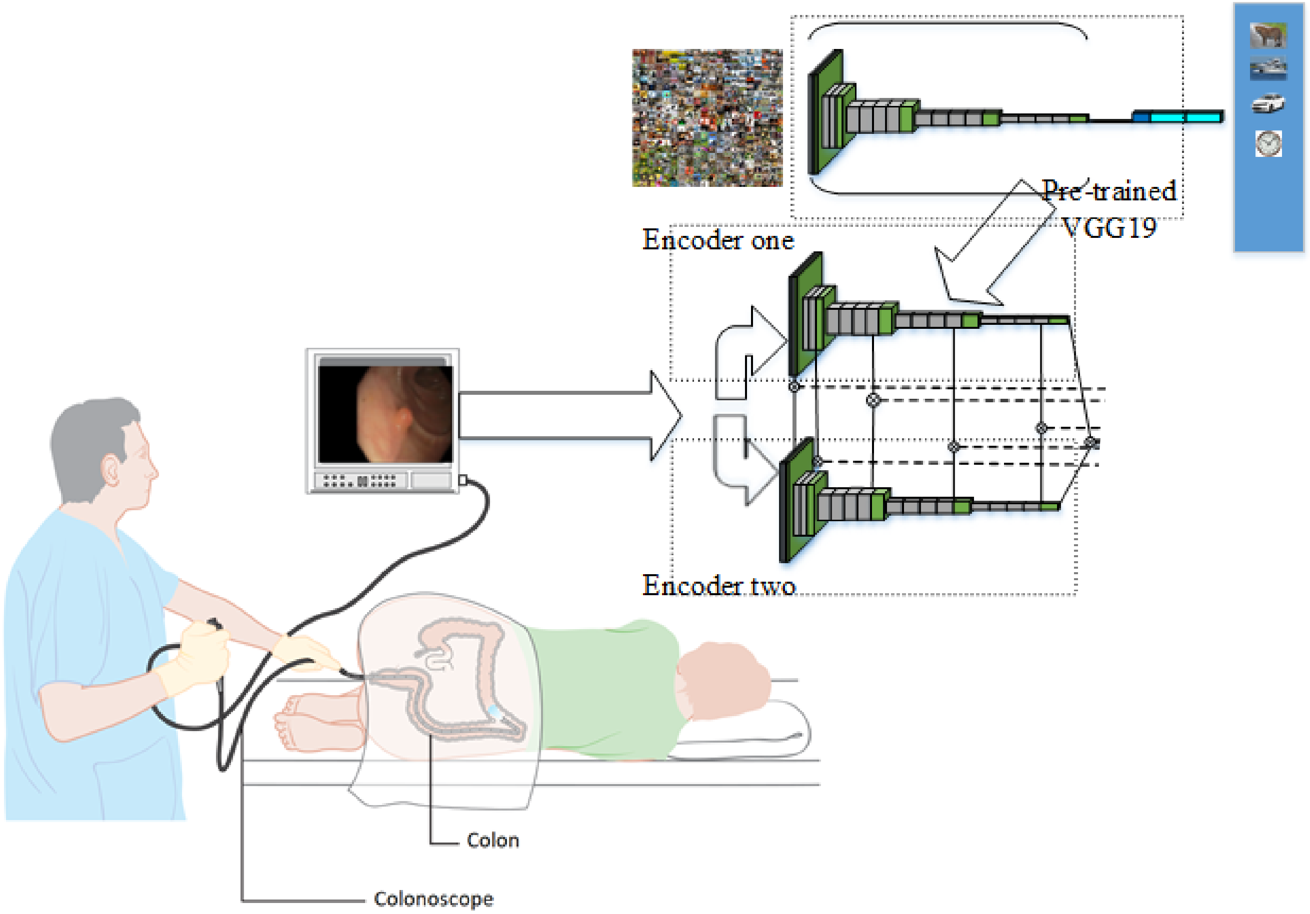


Proposed: Y-Net

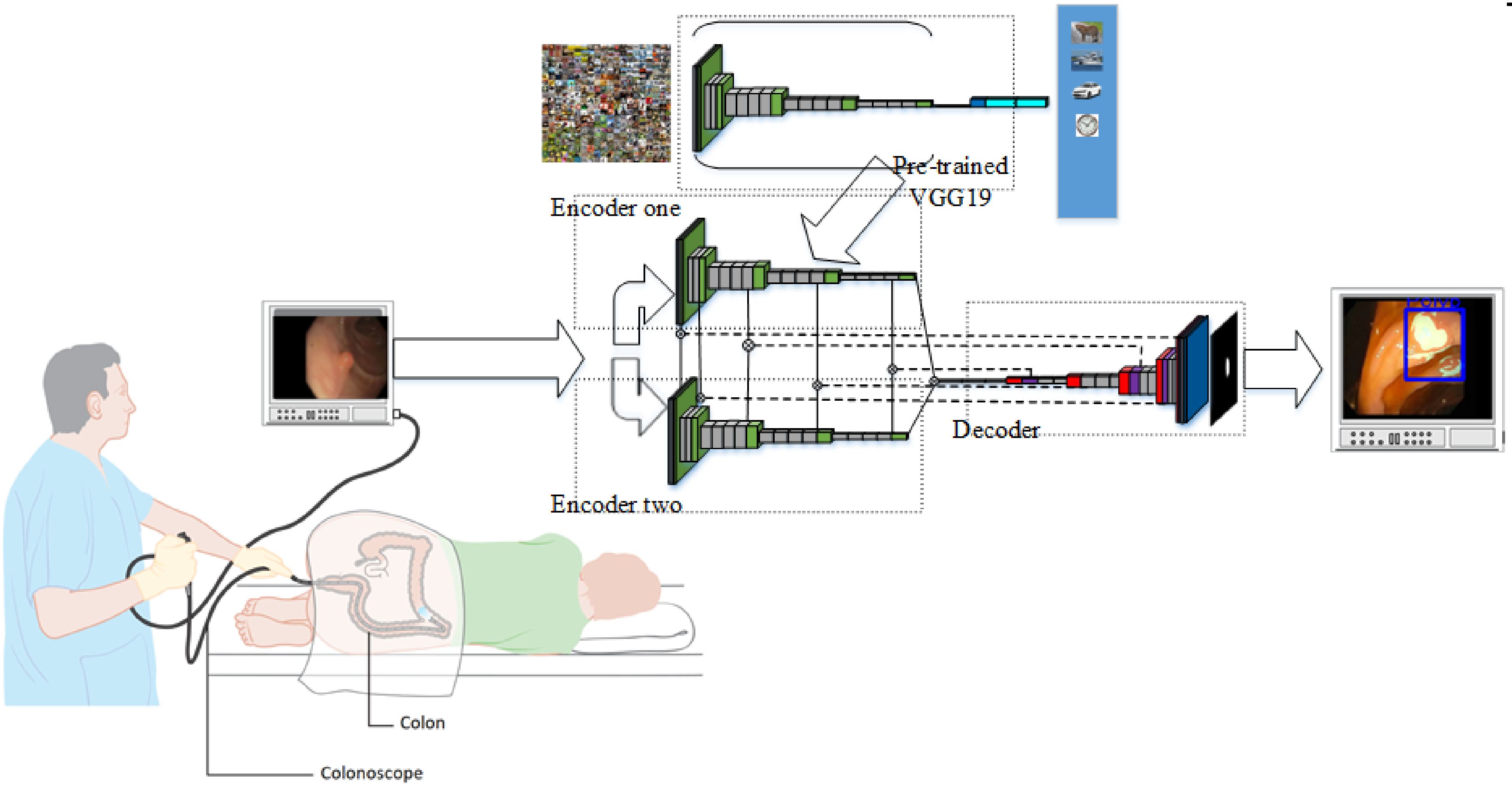




Proposed Method: Y-Net



Proposed Method: Y-Net



Training

$$\theta_{t+1} = \theta_t - \frac{c \cdot \eta}{\sqrt{E[g^2] + \epsilon}} \cdot g_t$$

$$\mathcal{L}(p, g) = -\frac{1}{N} \sum_{i=1}^{N} \left(\frac{\lambda}{2} \cdot g_i \cdot \log p_i \right) + \left(1 - \frac{2 \sum_{i=1}^{N} (g_i \cdot p_i) + \epsilon}{\sum_{i=1}^{N} (p_i) + \sum_{i=1}^{N} (g_i) + \epsilon} \right)$$

Dataset:

- ASU-Mayo clinic polyp database
- 20 and 18 short segment colonoscopy videos for training and testing with pixel level annotated polyp masks
- 4278/18495 frames with polyps in the training set and 4300/17574 frames in test set.

Evaluation metrics:

- F1 and F2 Metrics

$$P = \frac{N_{tp}}{N_{tp} + N_{fp}}, R = \frac{N_{tp}}{N_{tp} + N_{fn}}$$

$$F1 = \frac{2PR}{P + R}, F2 = \frac{5PR}{4P + R}$$

Ablation study

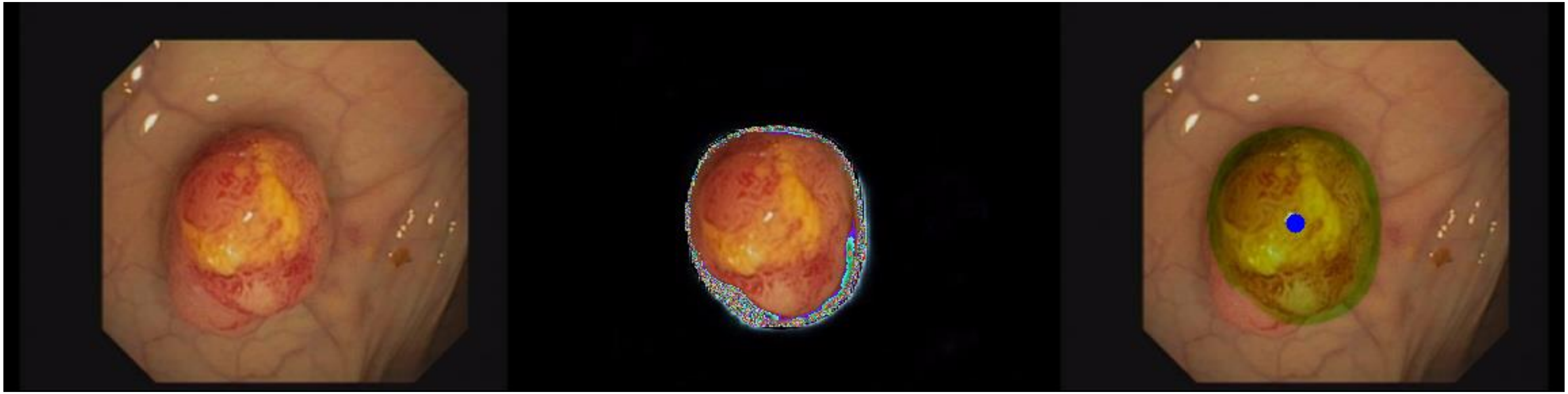
Method	Prec[%]	Rec[%]	F1[%]	F2[%]
U-Net(Trained from scratch, baseline)	90.8	39.2	54.7	44.2
U-Net(Pre-trained encoder VGG19, single encoder)	**96.2**	68.2	79.8	72.4
Y-Net(Ours)	87.4	**84.4**	**85.9**	**85.0**

Result: Comparison

Method	TP	FP	FN	Prec[%]	Rec[%]	F1[%]	F2[%]
PLS	1594	10103	2719	13.6	36.9	19.9	27.5
CVC-CLINIC [1]	1578	3456	2735	31.3	36.6	33.8	35.4
OUS	2222	229	2091	90.6	51.5	65.7	56.4
ASU [2]	2636	184	1677	**93.5**	61.1	73.9	65.7
CUMED	3081	769	1232	80.0	71.4	75.5	73.0
Fusion [3]	3062	414	1251	88.1	71.0	78.6	73.9
Y-Net(Ours)	3582	513	662	87.4	**84.4**	**85.9**	**85.0**

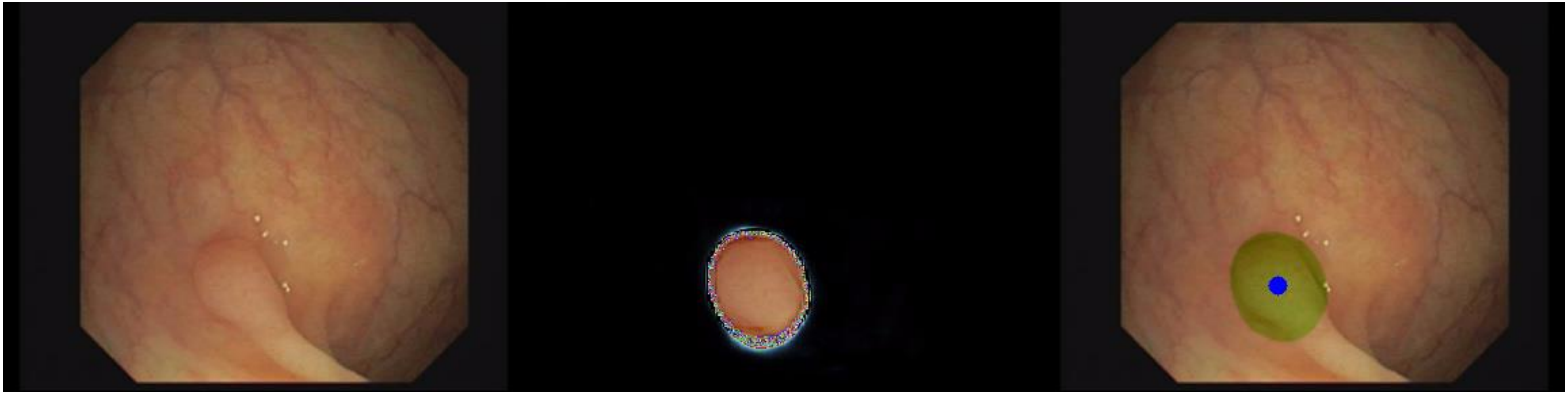
1. Bernal et al., "WM-DOVA maps for accurate polyp highlighting in colonoscopy: Validation vs. saliency maps from physicians".
2. Tajbakhsh, Gurudu, and Liang, "Automated polyp detection in colonoscopy videos using shape and context information".
3. Yu et al., "Integrating online and offline three-dimensional deep learning for automated polyp detection in colonoscopy videos".

Result:



Mohammed, A., Yildirim, S., Farup, I., Pedersen, M. and Hovde, Ø., 2018. Y-net: A deep convolutional neural network for polyp detection. *arXiv preprint arXiv:1806.01907*.

Result:



Mohammed, A., Yildirim, S., Farup, I., Pedersen, M. and Hovde, Ø., 2018. Y-net: A deep convolutional neural network for polyp detection. *arXiv preprint arXiv:1806.01907*.

PS-DeVCEM: Pathology-sensitive deep learning model for video capsule endoscopy based on weakly labeled data.

- Limited amount of annotated samples
- ≈50,000 images
- Manually labeling images for pathologies simply does not scale well and is expert-intensive.

MIL: Multiple instance learning

- A type of weakly supervised learning problem where only group-level, also known as bag level annotation, is available. The instances within the bag are not labeled.

  E.g: the annotation could be a general statement about the category of the pathology in the video without information about the location within the video or frame labels.

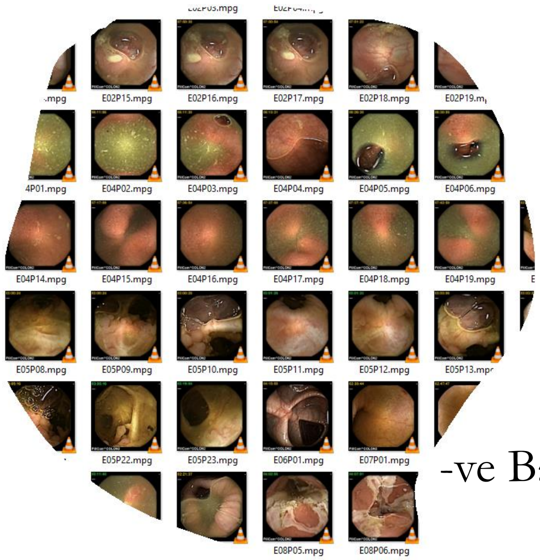
- **Independent samples (images):**
  - E.g. Single-instance learning (SIL): assigns each instance the label of its bag, creating a supervised learning problem, but mislabeling negative instances in positive bags [1]
- **Temporal based (video) MIL:**
  - E.g. The group-level prediction is given by taking the average of the instances. An objective function is introduced to encourage smoothness of inferred instance-level labels based on instance-level similarity, while at the same time respecting group-level label constraints. [2]

1. Doran, Gary, and Soumya Ray. "A theoretical and empirical analysis of support vector machine methods for multiple-instance classification." *Machine learning 97*, no. 1-2 (2014): 79-102.
2. Kotzias, D., Denil, M., De Freitas, N. and Smyth, P., 2015, August. From group to individual labels using deep features. *In Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (pp. 597-606).

Assumptions

- It is assumed that positive bag videos contain at least one instance of a given pathology while a negative bag video depicts none.

+ve Bag



-ve Bag

Let $V = \{f_1, f_2, f_3,\ldots, f_N\}$ be a video containing frames $f_1, f_2, f_3,\ldots, f_N$ and N is the number of frames in the video. We assume individual labels are available for each video V and is given by G with unknown frame label $\boldsymbol{y} = \{y_1, y_2, y_3,\ldots, y_N\}$.

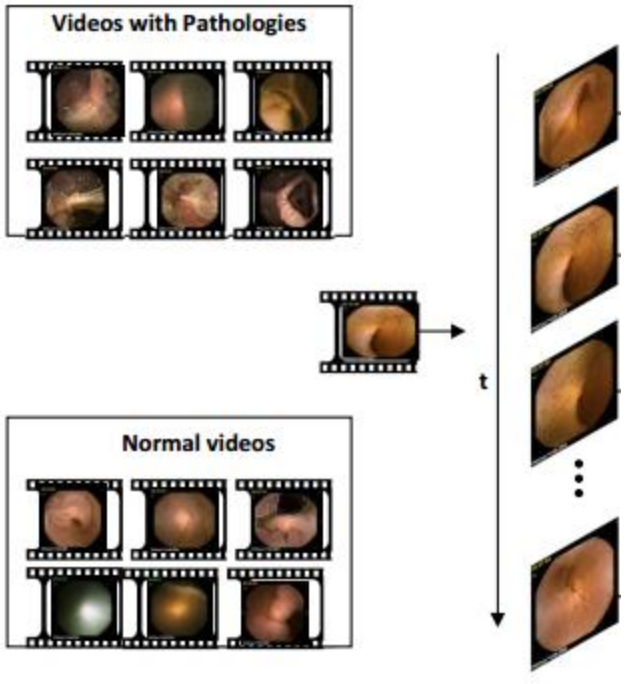
MIL constraints

$$Y = \begin{cases} p & \text{if } \exists n \quad s.t. \quad y_n = p, p \subseteq P, n \in N \\ 0, & \text{otherwise} \end{cases}$$

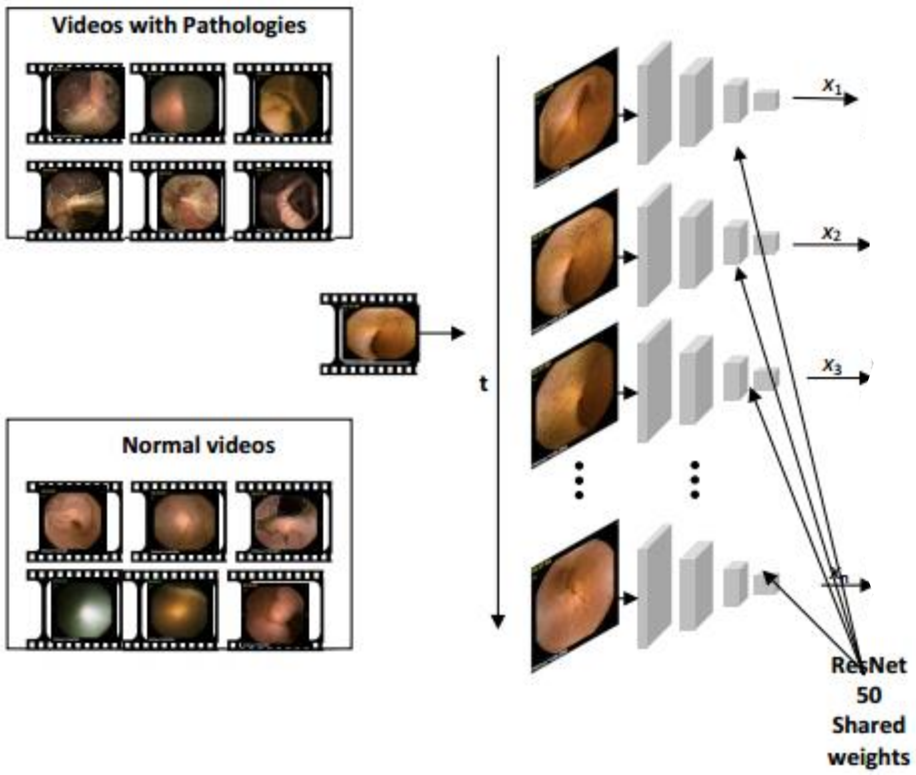
Alternative MIL constraint

$$Y = \max_n \{y_n\} \mid \quad y_n = p, p \subseteq P, n \in N$$

PS-DeVCEM

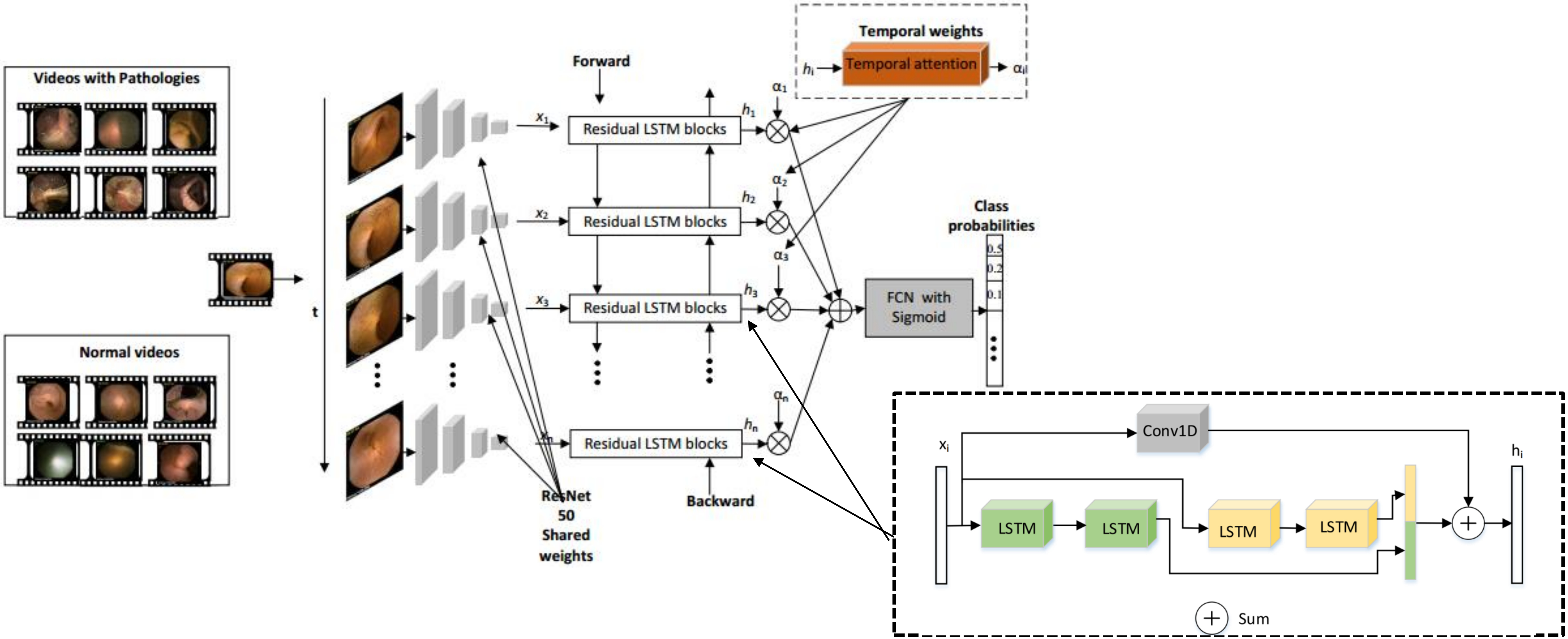


PS-DeVCEM



Videos with Pathologies
Normal videos
t
x_1
x_2
x_3
x_n
ResNet
50
Shared
weights

PS-DeVCEM

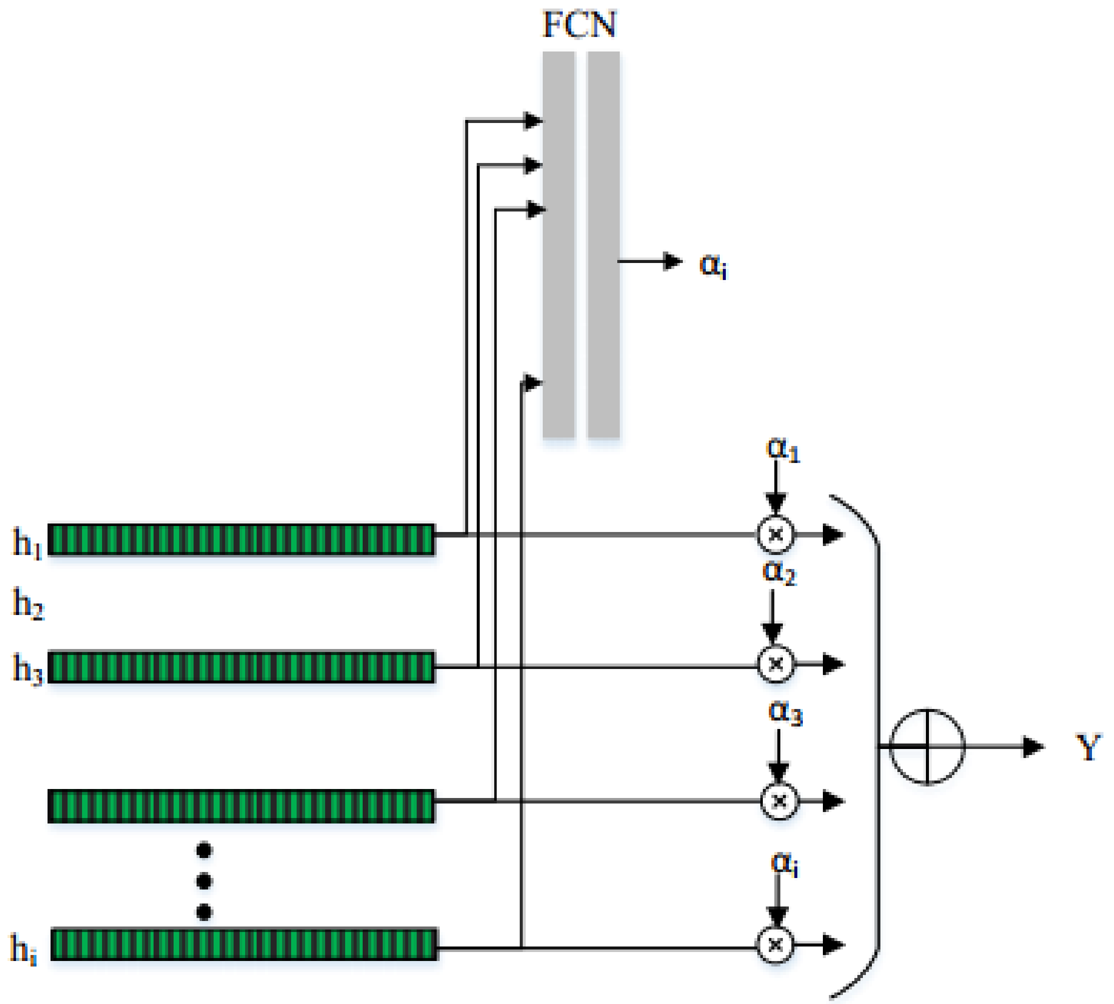

Videos with Pathologies
Normal videos
t
Forward
Backward
Residual LSTM blocks
Residual LSTM blocks
Residual LSTM blocks
Residual LSTM blocks
ResNet 50 Shared weights
Temporal weights
Temporal attention
Class probabilities
FCN with Sigmoid
0.5
0.2
0.1
Conv1D
LSTM
LSTM
LSTM
LSTM
Sum

Attention

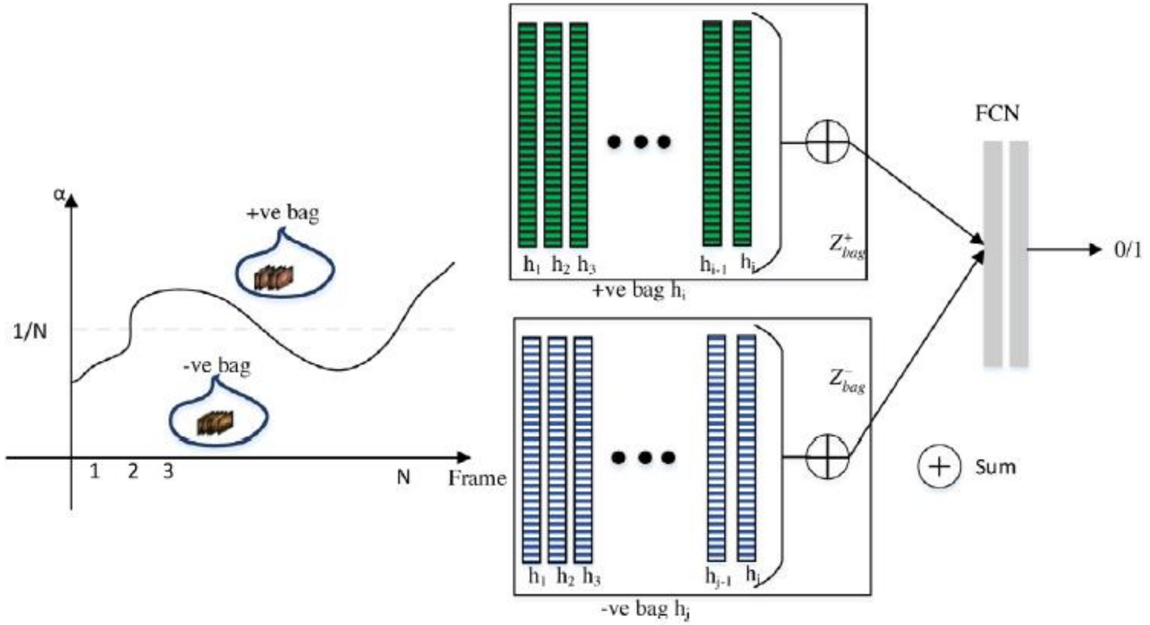
- Provides insight into the contribution of each instance to the bag label.

$$Z = \sum_{n=1}^{N} \alpha_n h_n$$

$$\alpha_n = \frac{exp\{\boldsymbol{w}^T \tanh(\boldsymbol{V}\boldsymbol{h}_n^T)\}}{\sum_{i=1}^{N} exp\{\boldsymbol{w}^T \tanh(\boldsymbol{V}\boldsymbol{h}_i^T)\}}$$



Training



$$Z^+_{bag} = \sum_{b=1}^{B^+=|\alpha > \frac{1}{N}|} h^+_b \mid h^+_b \subseteq \boldsymbol{h}, \alpha_b > \frac{1}{N}$$

$$Z^-_{bag} = \sum_{b=1}^{B^-=|\alpha \leq \frac{1}{N}|} h^-_b \mid h^-_b \subseteq \boldsymbol{h}, \alpha_b \leq \frac{1}{N}$$

$$\mathcal{L} = \frac{1}{M}\sum_{m=1}^{M} g_n \log(y_n) +$$

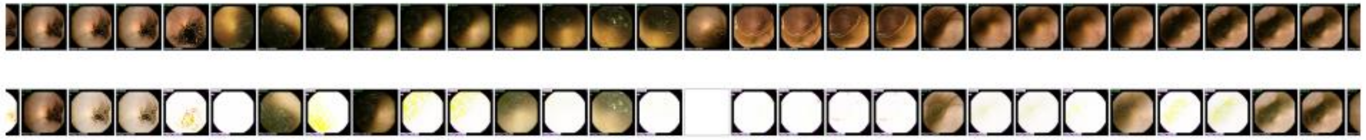
$$-\frac{\lambda}{M}\sum_{m=1}^{M} g_n^{bag} \log(y_n^{bag}) - (1 - g_n^{bag}) \log(1 - y_n^{bag})$$

Annotation:

Pathology	Erosions	Debris	Diverticulosis	Erythema	Granularity	Hemorrhage	Inflammation	Normal	Edema	Angioectasia	Polyp	Pseudopolyp	Tumor	Ulceration	**Total**
# training videos	54	72	17	16	27	17	22	45	5	1	32	28	8	32	**227**
# testing videos	64	84	16	21	28	20	24	41	7	1	30	29	3	45	**228**

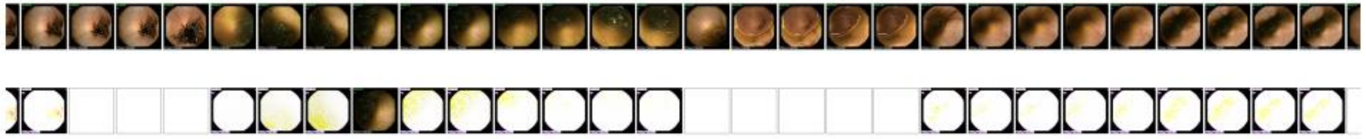
Ablation study

Train: Each convolution feature is weighted with computed value α_n before feeding into the LSTM network.



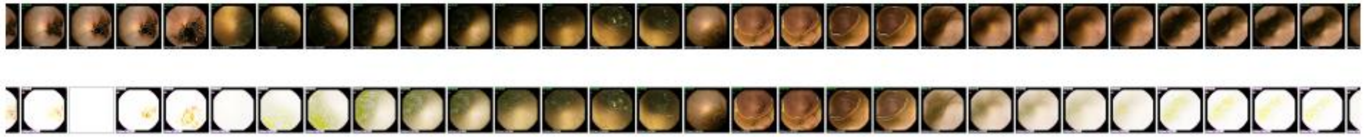
(a) AttenConv attention. Predicted: 'Debris', 'Erosion'

Input: Frame feature
Attend: Frame feature



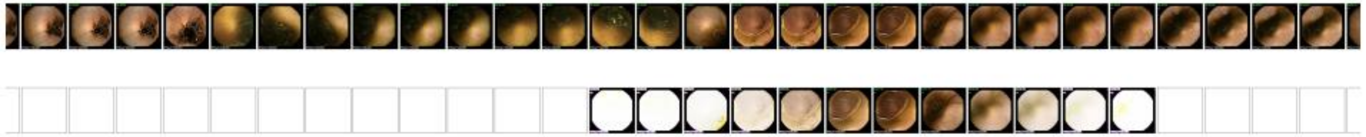
(b) AttenConvLSTM attention. Predicted: 'Polyp', 'Debris'

Input: Frame feature
Attend: Hidden state of LSTM block



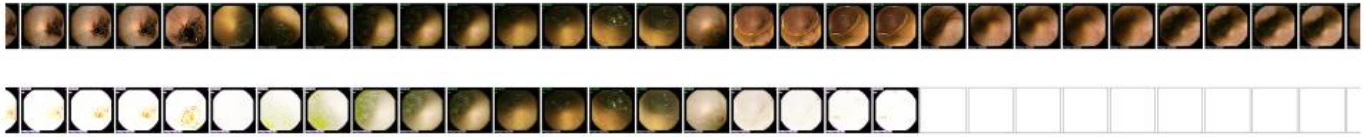
(c) AttenLSTM attention. Predicted: 'Polyp', 'Debris'

Input: Hidden state of LSTM block
Attend: Hidden state of LSTM block



(d) GiudedLSTM attention. Predicted: 'Polyp', 'Debris'

Input: Hidden state of LSTM block
Attend: Hidden state of LSTM block + self-supervision



(e) EndoscopicMIL(proposed) attention. Predicted: 'Polyp', 'Debris'

Input: Residual LSTM block
Attend: Residual LSTM block + self-supervision

Ablation study

Ablation study result: The values are averaged for all pathologies.

Method	Precision	Recall	F1-score	Specificity
AttenConv	0.229	0.290	0.246	0.872
AttenConvLSTM	0.450	0.461	0.443	0.939
AttenLSTM	0.529	0.478	0.487	**0.954**
GuidedLSTM	0.487	0.482	0.458	0.946
EndoscopicMIL(proposed)	**0.616**	**0.546**	**0.551**	0.951

Result

Comparison with other MIL algorithms: The values are averaged for all pathologies.

Method	Precision	Recall	F1-score	Specificity
SIL [1]	0.235	0.046	0.066	**0.997**
MissSVM [2]	0.130	0.162	0.123	0.912
Attention based deep MIL [3]	**0.616**	0.471	0.513	0.955
STPN [4]	0.592	0.517	0.536	0.916
W-TALC [5]	0.274	**0.891**	0.416	0.666
EndoscopicMIL(w/o self-supervision)	0.606	0.54	0.54	0.951
EndoscopicMIL(proposed)	**0.616**	0.546	**0.551**	0.951

1. S. Ray, M. Craven, Supervised versus multiple instance learning: An empirical comparison, in: Proceedings of the 22nd international conference on Machine learning, ACM, 2005, pp. 697–704.
2. Z.-H. Zhou, J.-M. Xu, On the relation between multi-instance learning and semi-supervised learning, in: Proceedings of the 24th international conference on Machine learning, ACM, 2007, pp. 1167–1174.
3. M. Ilse, J. M. Tomczak, M. Welling, Attention-based deep multiple instance learning, 2018, arXiv preprint arXiv:1802.04712.
4. L. Wang, Y. Xiong, D. Lin, L. Van Gool, Untrimmednets for weakly supervised action recognition and detection, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2017, pp. 4325–4334
5. S. Paul, S. Roy, A. K. Roy-Chowdhury, W-talc: Weakly-supervised temporal activity localization and classification, in: Proceedings of the European Conference on Computer Vision (ECCV), 2018, pp. 563–579.

Visual result

Mohammed, A., Farup, I., Pedersen, M., Yildirim, S. and Hovde, Ø., 2020. PS-DeVCEM: Pathology-sensitive deep learning model for video capsule endoscopy based on weakly labeled data. *Computer Vision and Image Understanding*, *201*, p.103062.

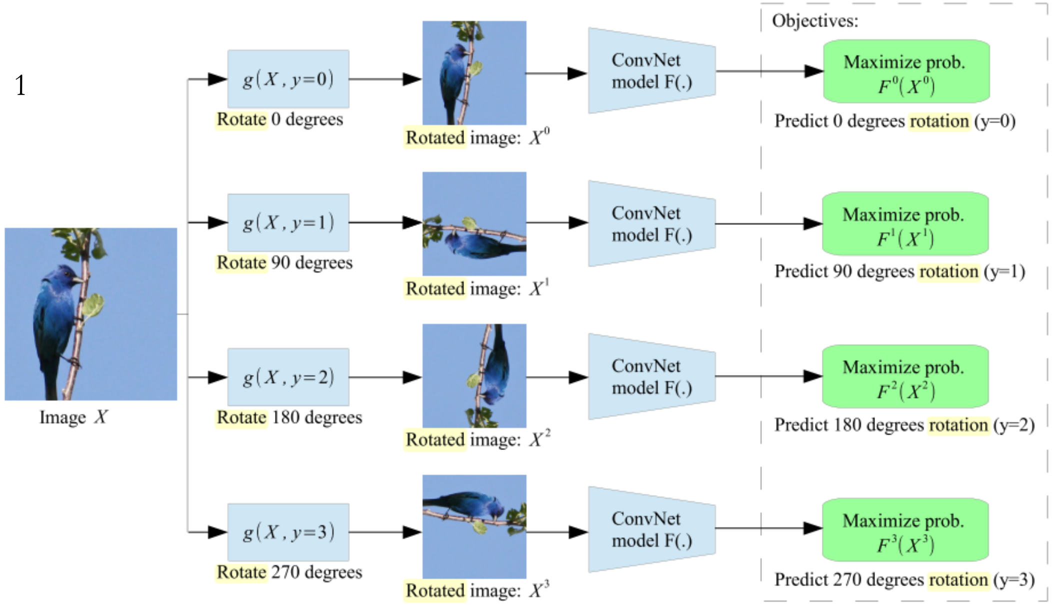
Visual result

Mohammed, A., Farup, I., Pedersen, M., Yildirim, S. and Hovde, Ø., 2020. PS-DeVCEM: Pathology-sensitive deep learning model for video capsule endoscopy based on weakly labeled data. *Computer Vision and Image Understanding*, *201*, p.103062.

Self-supervision:

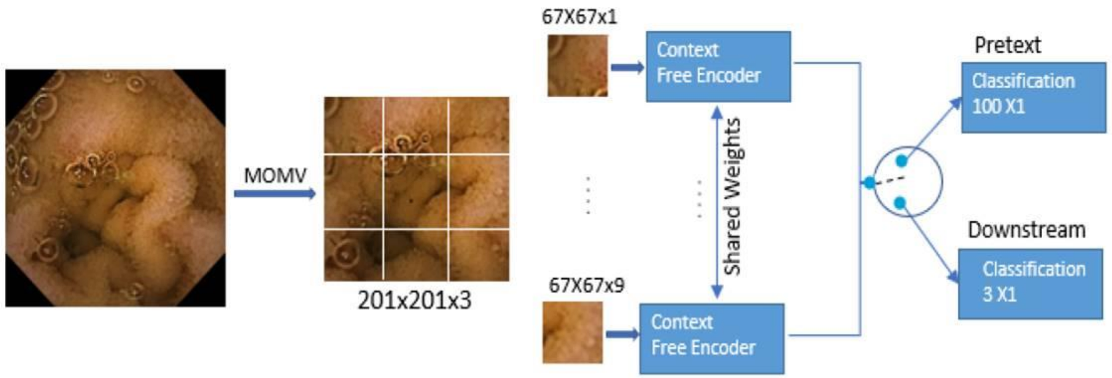
- Labels for free and train supervised
- Common in NLP

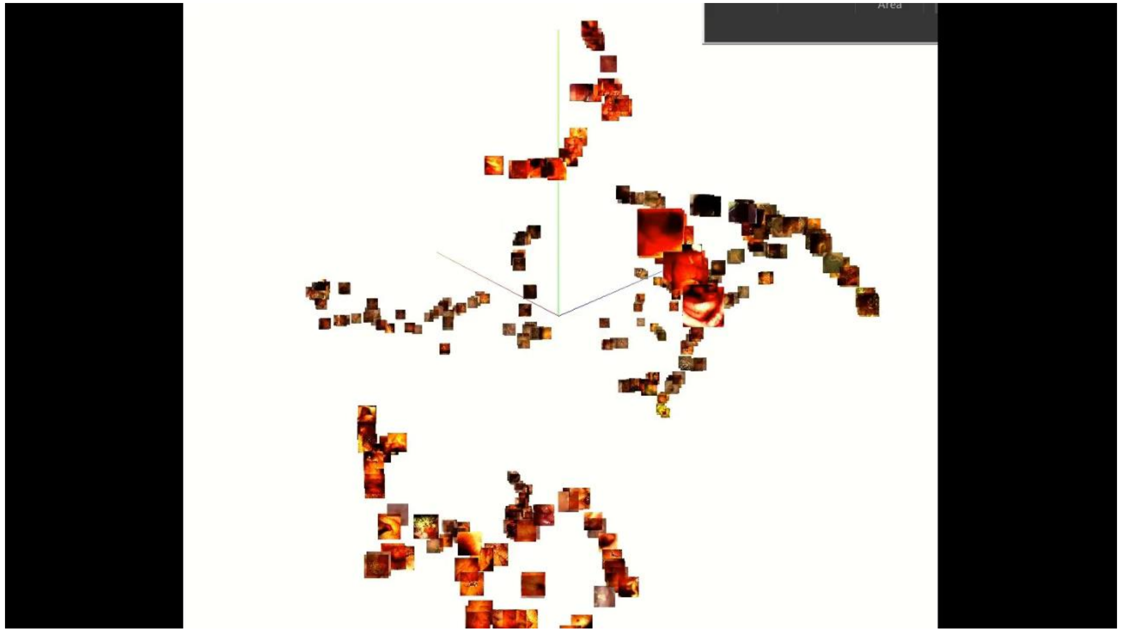
E.g. [1]

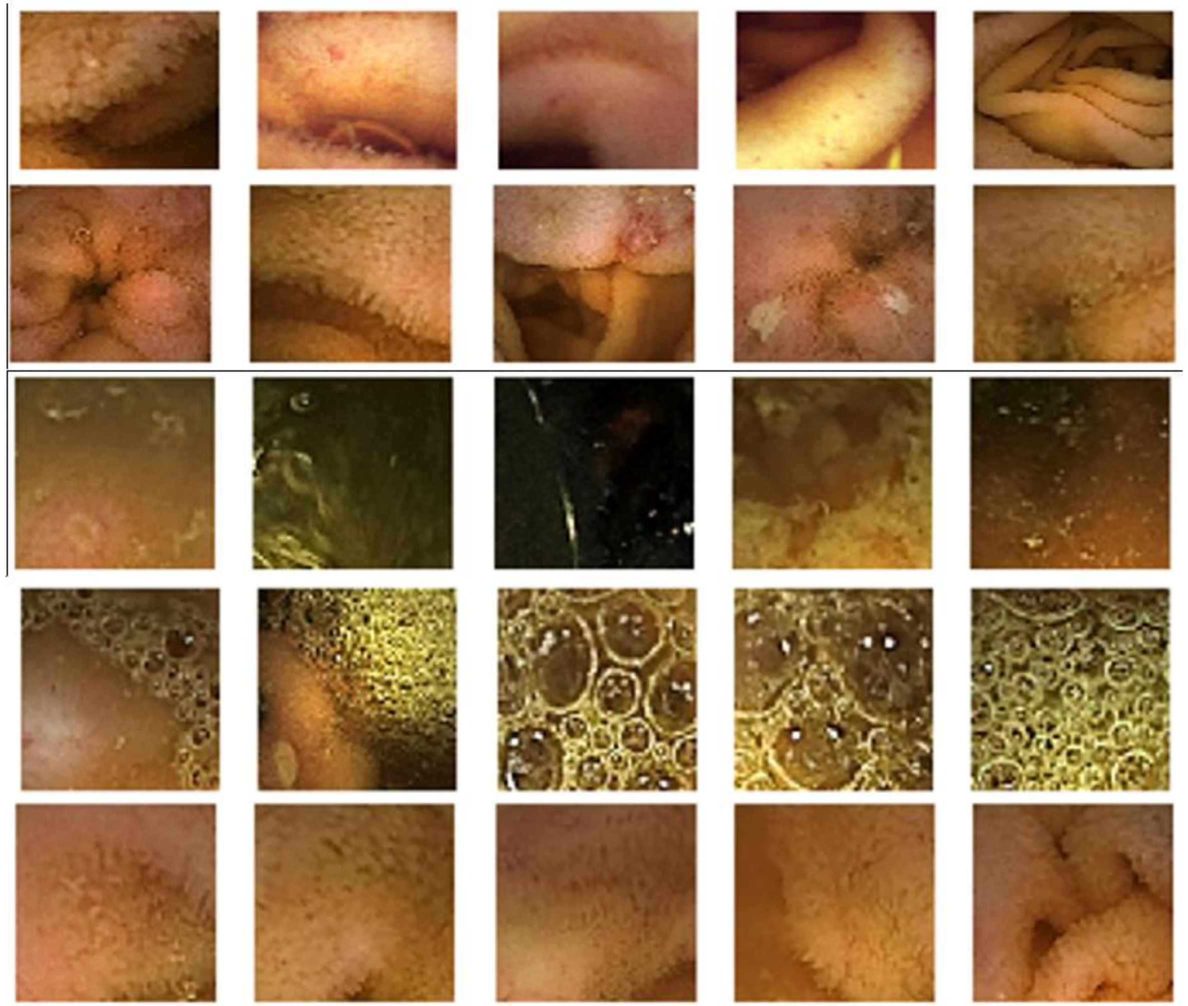


Gidaris, S., Singh, P. and Komodakis, N., 2018. Unsupervised representation learning by predicting image rotations. arXiv preprint arXiv:1803.07728.

Self-supervision



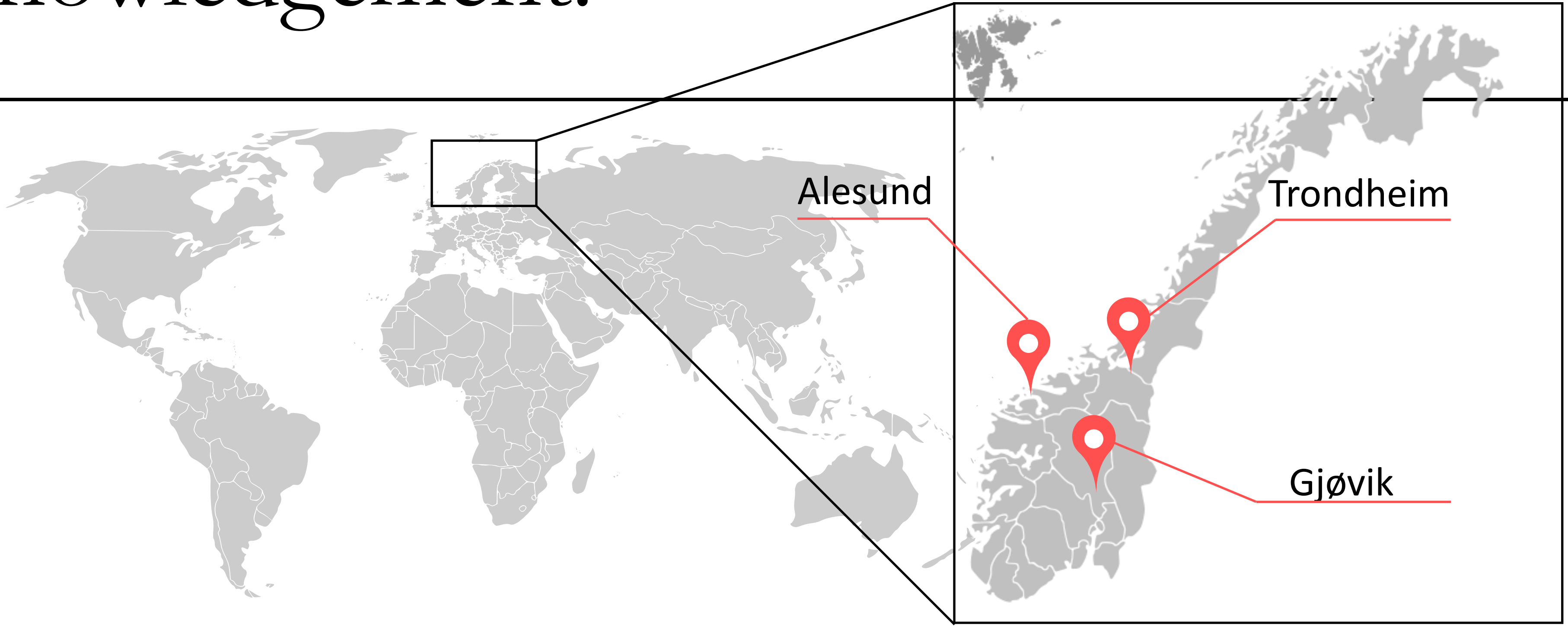




Conclusions and Future Work

- Improve the video frame localization through domain knowledge of the pathologies
- Weakly supervised video segmentation
- Novel self-supervision pre-text task/contrastive learning
- Explainability
- Advance to the next level and test these methods in real procedures.

Acknowledgement:



- Norwegian Colour and Visual Computing Laboratory
- “IQ-MED: Image Quality enhancement in MEDical diagnosis, monitoring and treatment, project no. 247689” and “CAPSULEAI3D Improved Pathology Detection in Wireless Capsule Endoscopy Images through Artificial Intelligence and 3D Reconstruction, project no. 300031