Polyp Detection Problems
October, 1st 2020

PREVIOUSLY ON…

Contents lists available at ScienceDirect

Computerized Medical Imaging and Graphics

journal homepage: www.elsevier.com/locate/compmedimag

WCE Polyp Detection with Triplet based Embeddings

Pablo Laiz[a,*], Jordi Vitrià[a], Hagen Wenzek[b], Carolina Malagelada[c], Fernando Azpiroz[c], Santi Seguí[a]

[a] *Department of Mathematics and Computer Science, Universitat de Barcelona, Barcelona, Spain*
[b] *CorporateHealth International ApS*
[c] *Digestive System Research Unit, University Hospital Vall d'Hebron, Barcelona, Spain*

ARTICLE INFO

Article history:

Keywords: Deep metric learning, Triplet loss, Deep learning, Capsule endoscopy, Polyp detection

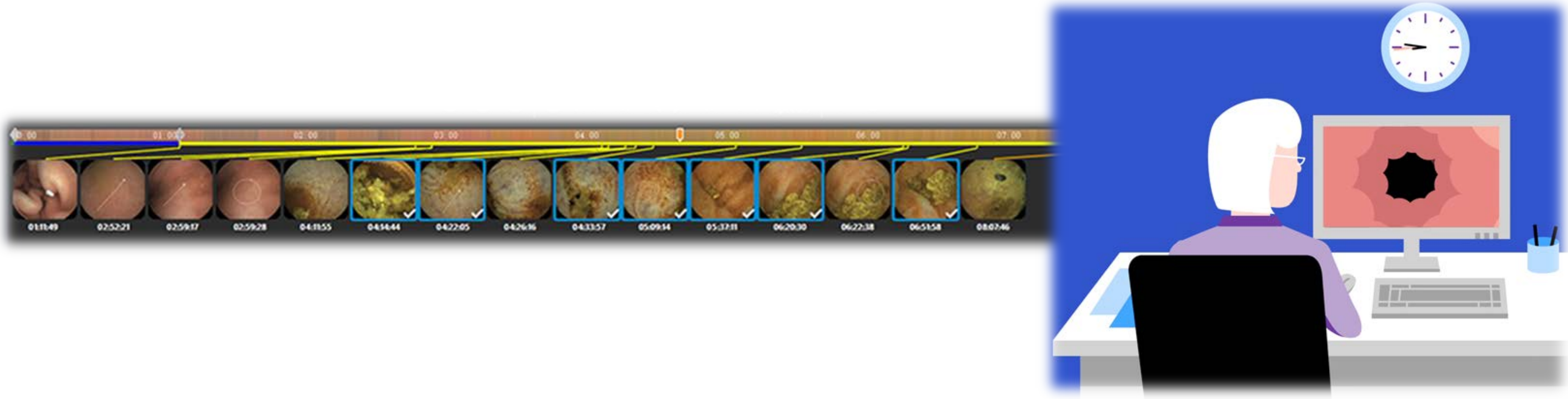
ABSTRACT

Wireless capsule endoscopy is a medical procedure used to visualize the entire gastrointestinal tract and to diagnose intestinal conditions, such as polyps or bleeding. Current analyses are performed by manually inspecting nearly each one of the frames of the video, a tedious and error-prone task. Automatic image analysis methods can be used to reduce the time needed for physicians to evaluate a capsule endoscopy video. However these methods are still in a research phase.

In this paper we focus on computer-aided polyp detection in capsule endoscopy im-

PROBLEM DEFINITION

- Identify the frames that may contain polyps by means of Capsule Endoscopy (CE).



OUR DATABASE

- What do we have?

120
procedures

1.3M
revised frames

2.1K
polyp frames

IMBALANCED DATASET

OUR DATABASE - POLYPS

- Only 52 out of the 120 procedures contain at least one polyp.
- A total of 165 polyps are found in these cases.
  - Each polyp may appear in various number of frames.
  - Each polyp has a different morphology and size.

# Frames	1-2	3-4	5-6	7-10	11-20	21+
# Polyps	33	32	20	19	31	30

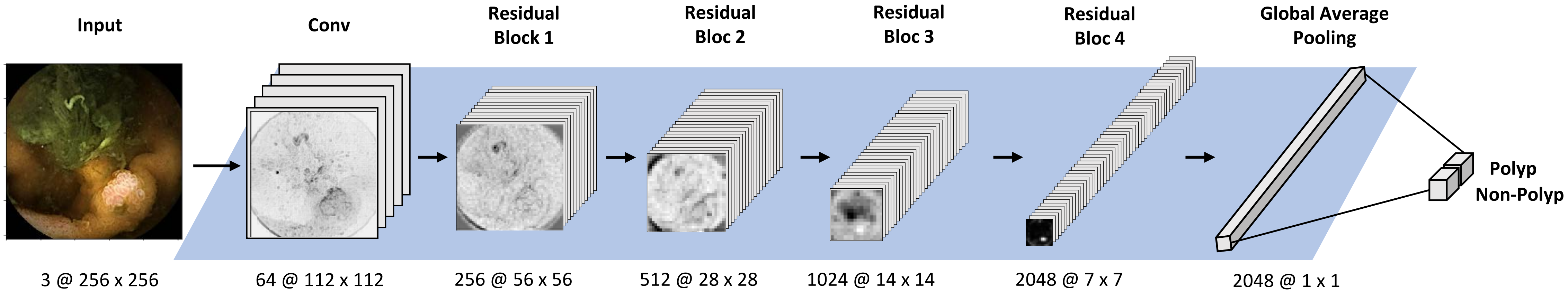
		Morphology			
		Sessile	**Pedunculated**	**Undefined**	**Total**
Size	**Small** (2-6 mm)	65	4	19	88
	Medium (7-11 mm)	29	4	20	53
	Large (12+ mm)	8	3	13	24
	Total	102	11	52	165

HYPOTHESIS

Triplet loss might help to reduce the impact of a small imbalanced dataset over the CAD support system.

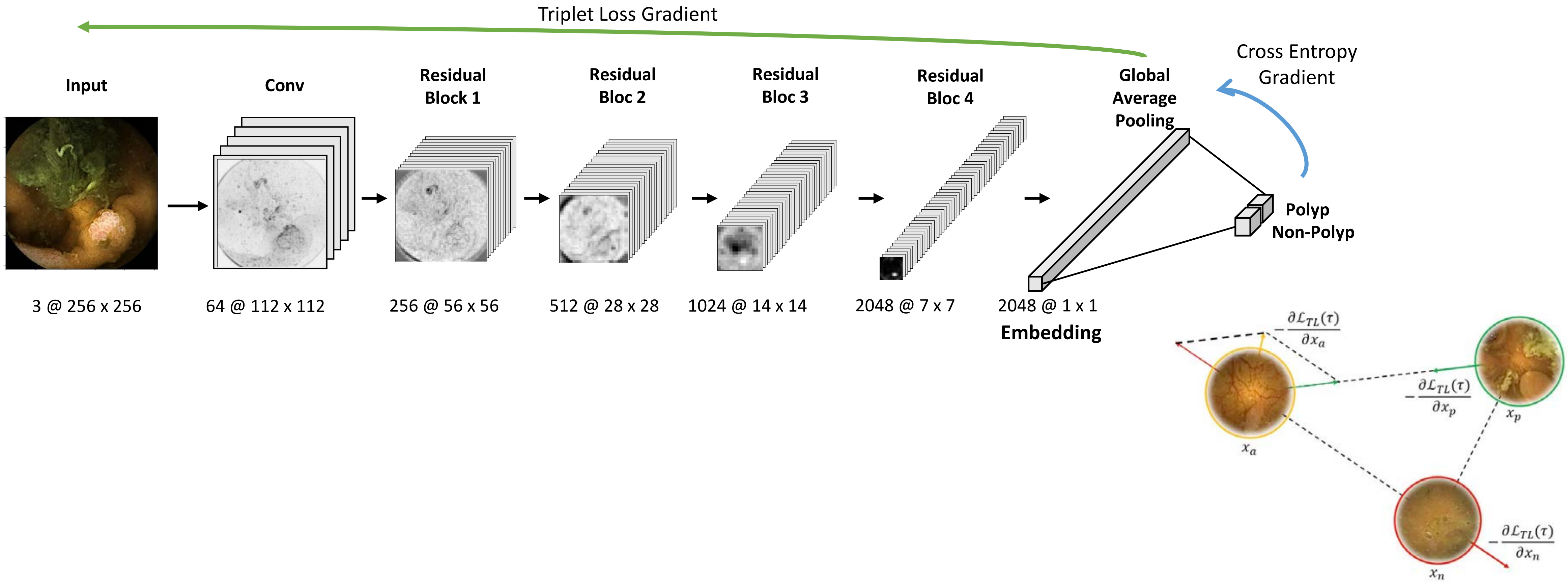
OUR PAPER APPROACH

- ResNet 50 + Transfer Learning (ImageNet)



OUR PAPER APPROACH

- Triplet loss to enhance the image representation.



OUR PAPER APPROACH

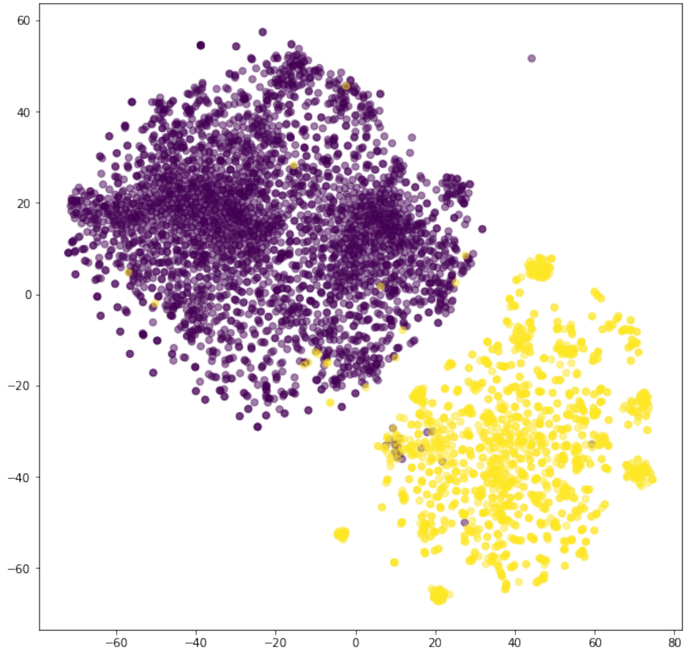
- Results

Parameter Optimization	Accuracy (%)	Sensitivity (%)	Specificity (%)	AUC (%)	Sensitivity (%)		
					Spec. at 95%	**Spec. at 90%**	**Spec. at 80%**
ResNet	97.85 ± 0.24	26.01 ± 9.78	97.97 ± 0.27	82.85 ± 5.72	37.75 ± 9.12	51.49 ± 11.09	66.71 ± 12.15
SSAEIM	59.85 ± 48.75	40.11 ± 48.90	59.91 ± 48.92	57.76 ± 5.83	6.98 ± 2.99	13.29 ± 3.56	27.82 ± 5.96
UDCS	94.41 ± 1.53	71.51 ± 7.80	94.45 ± 1.54	88.64 ± 2.87	70.44 ± 6.53	78.22 ± 6.46	83.31 ± 5.18
ANET	96.96 ± 0.53	**65.07 ± 7.58**	97.02 ± 0.54	90.44 ± 3.23	72.02 ± 6.03	78.92 ± 5.59	85.23 ± 4.98
TL_{BH}	**99.83 ± 0.05**	0.00 ± 0.00	**100.00 ± 0.00**	87.68 ± 2.71	50.15 ± 3.21	63.19 ± 4.48	77.52 ± 6.70
TL_{BA}	**99.43 ± 0.12**	51.15 ± 7.62	99.51 ± 0.12	**92.94 ± 1.87**	**76.68 ± 4.93**	**82.86 ± 4.78**	**88.53 ± 3.76**

But still there
are some issues....

GENERALIZATION PROBLEM

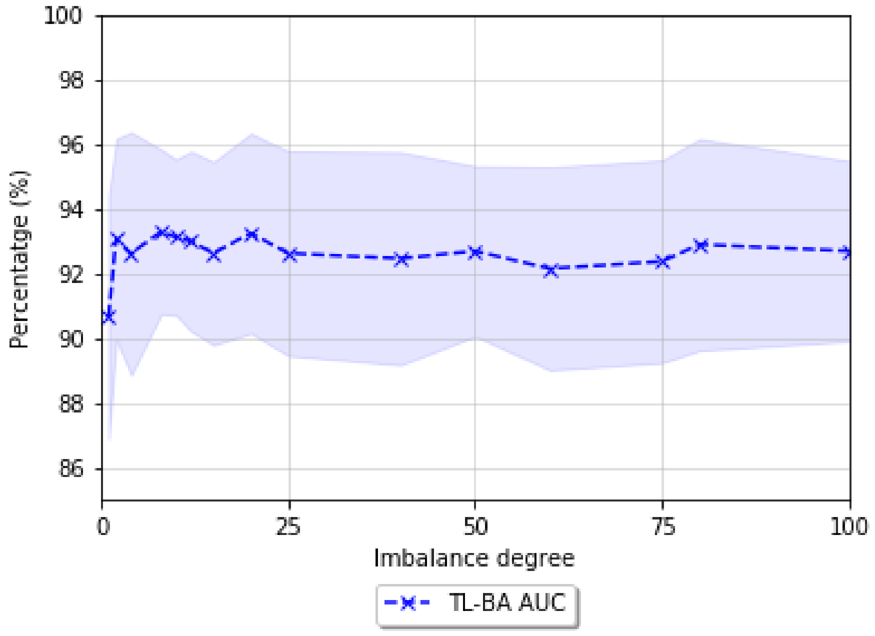
- Although the test results seems good, the system overfit with few epochs.



Train distribution

GENERALIZATION PROBLEM

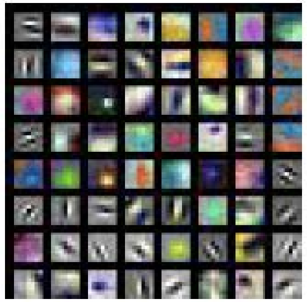
- It seems that the system does not achieve a appropriate local minima.
- The system performance can change dramatically from one run to another.



1) RESNET + TRANSFER LEARNING

- ImageNet has quite remarkable filters for visual object recognition.
- To detect GI pathologies, the networks need to find other patterns.

Dog sample ImageNet

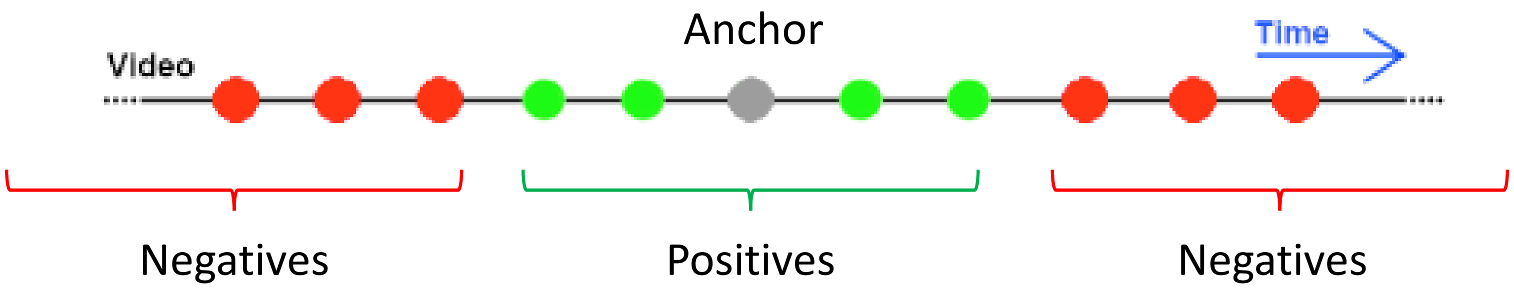


ImageNet Filters

Polyp sample

1) RESNET + TRANSFER LEARNING

- GOAL: Build a new initialization for the network to improve the performance of the system.
- It is used the Triplet loss to build unsupervised models to initialize polyp detection networks. Hence, pseudo labels are created considering the frame distance.



1) RESNET + TRANSFER LEARNING

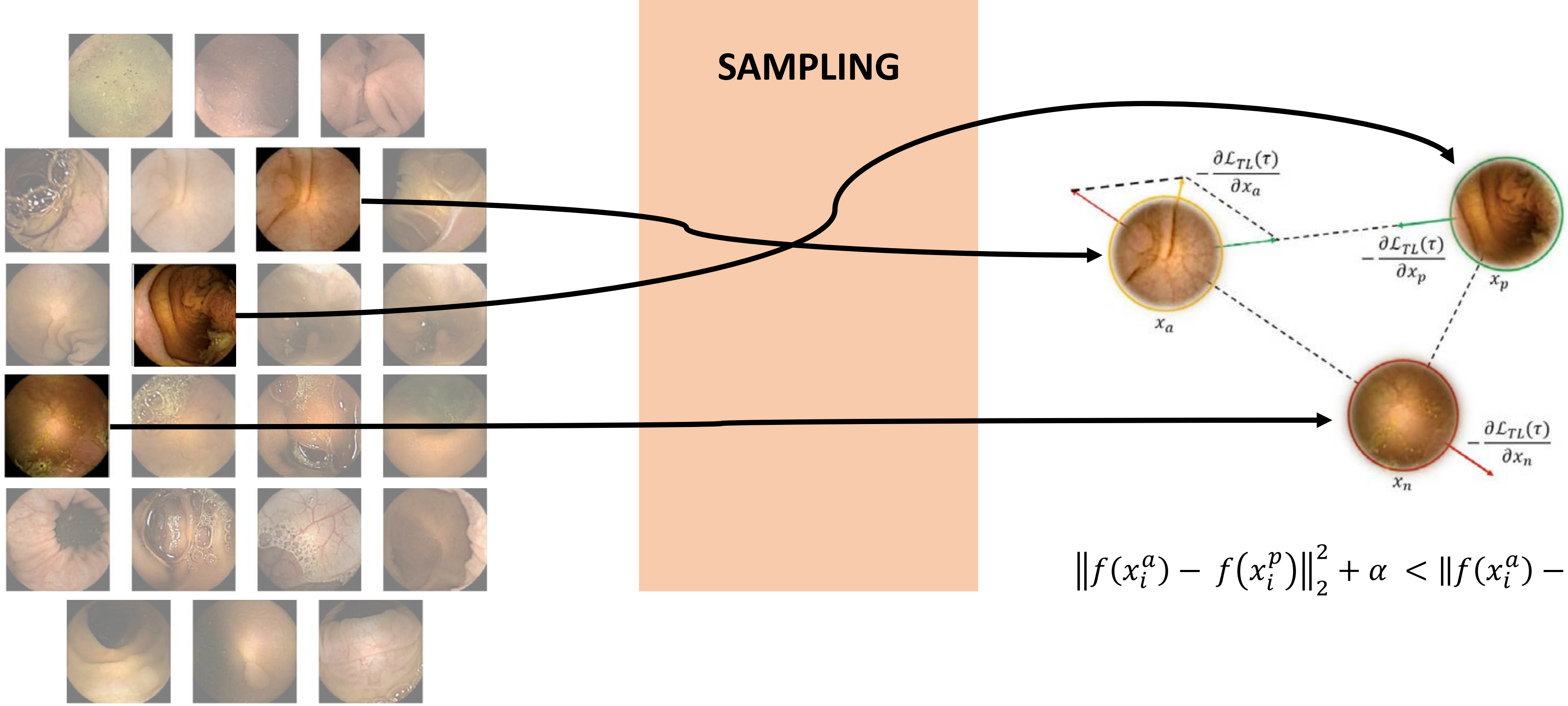
- Results

POLYP DETECTION RESULTS

Method	Sensitivity (%) Spec. at 95%	Spec. at 90%	Spec. at 80%	AUC (%)
ImageNet + [4]	76.68 ± 4.93	82.86 ± 4.78	88.53 ± 3.76	92.94 ± 1.87
Ours + [4]	$\mathbf{78.80 \pm 5.73}$	$\mathbf{85.17 \pm 4.61}$	$\mathbf{91.58 \pm 2.73}$	$\mathbf{94.99 \pm 1.27}$

2) TRIPLET LOSS

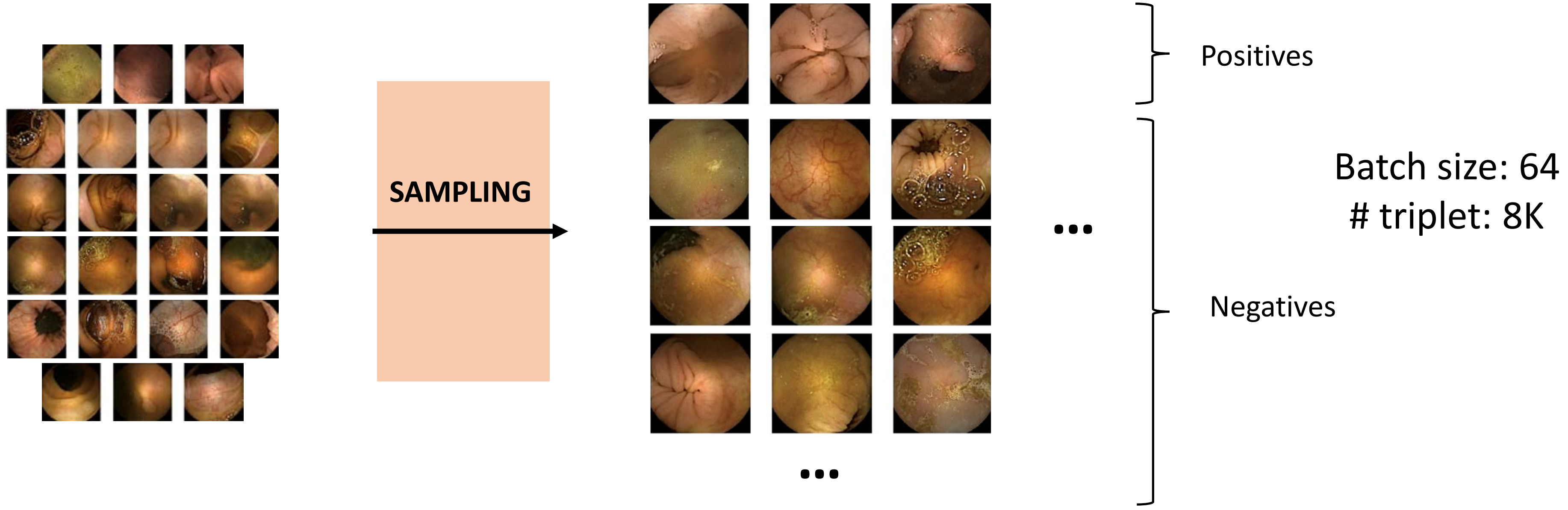
- In deep metric learning, the loss function is as important as the building of the batches.



$$\|f(x_i^a) - f(x_i^p)\|_2^2 + \alpha < \|f(x_i^a) - f(x_i^n)\|_2^2$$

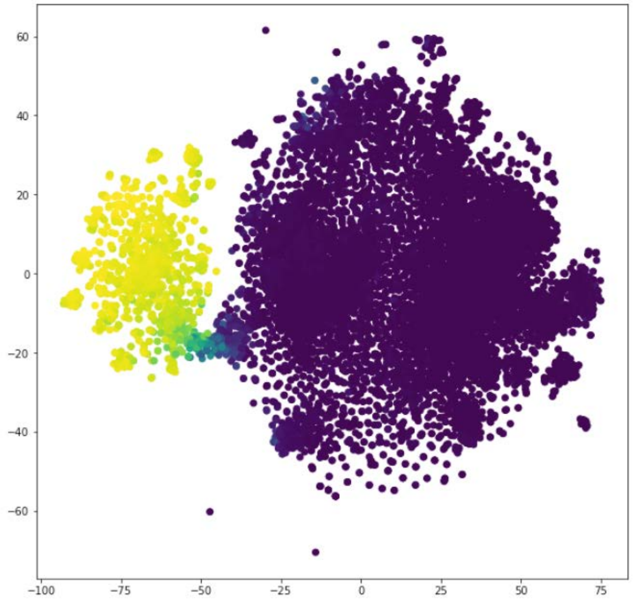
2) TRIPLET LOSS – SAMPLING MATTERS

- Current sampling:
  - Randomly selected images.
  - For each positive image, four negatives are sampled.
  - When all the positives are sampled, the epoch finishes.
    - All negatives frames are not chosen.



2) TRIPLET LOSS – SAMPLING MATTERS

- How to improve the sampling?
  - Using <u>meaningful negatives</u> that would add relevant information to the gradient:
    - Selecting wrong classified training data.
    - Selecting those negatives that are close to the positives in training data.
    - Selecting those images that are far from the cluster.



T-SNE with a sample of training data

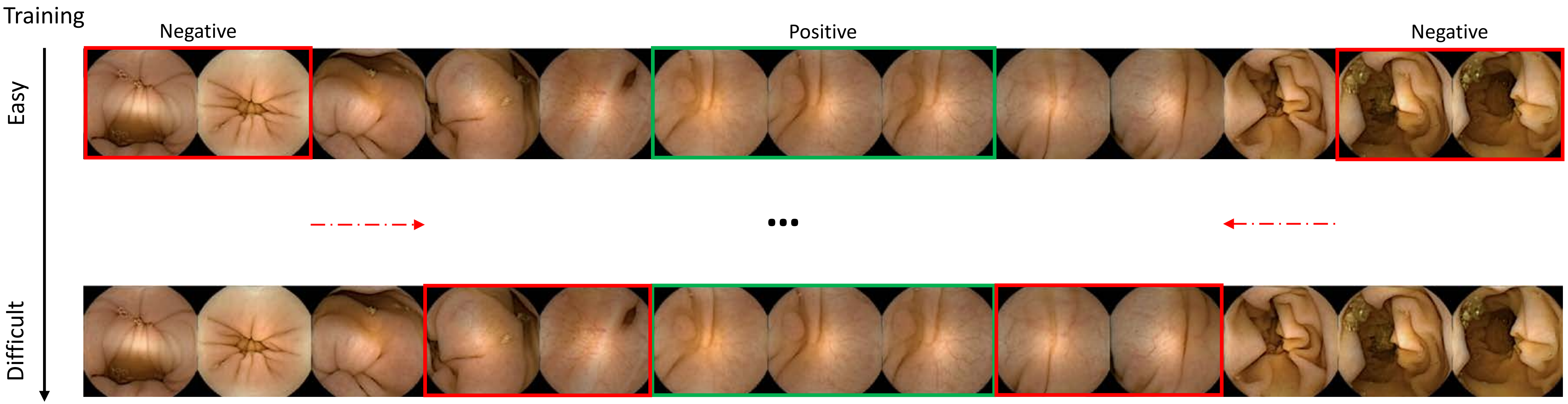
"Meaningful negatives = False positive"

2) TRIPLET LOSS – SAMPLING MATTERS

- How to improve the sampling?
  - Down sampling the negative set by removing similar images.
    - Currently DB is built from 100-frames sequences.
  - Each polyp appears in various number of frames:
    - Sampling by unique polyp, not by frame.
    - Same approach, but with morphology or size

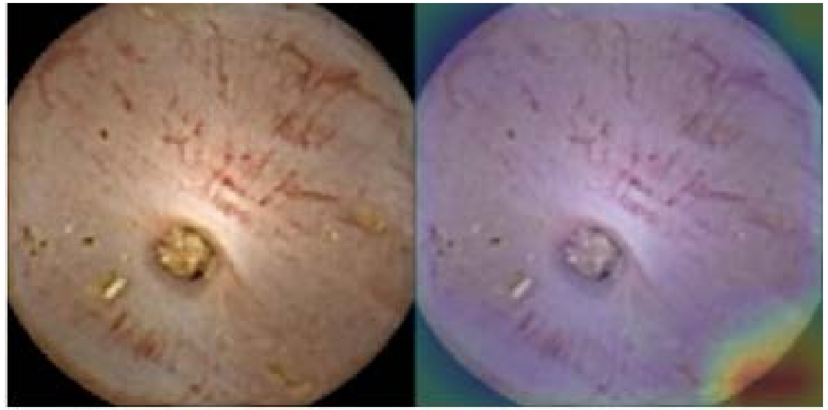
2) TRIPLET LOSS – SAMPLING MATTERS

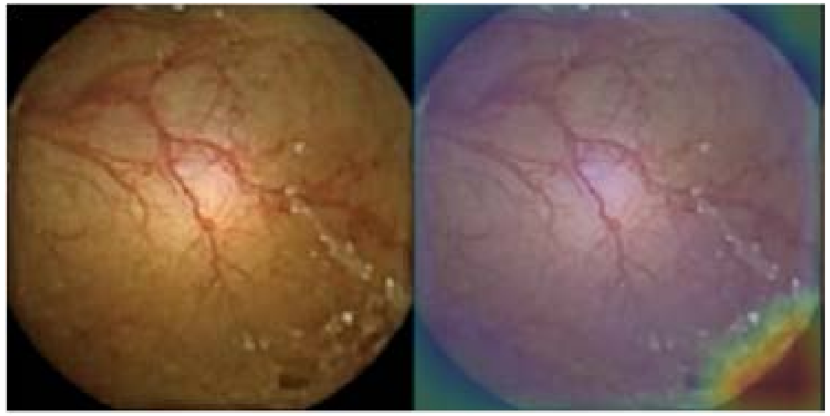
- How to improve the sampling?
  - Curriculum learning:
    - Start using far negatives frames from the positive and each epoch takes nearest samples.



BLACK REGIONS PROBLEM

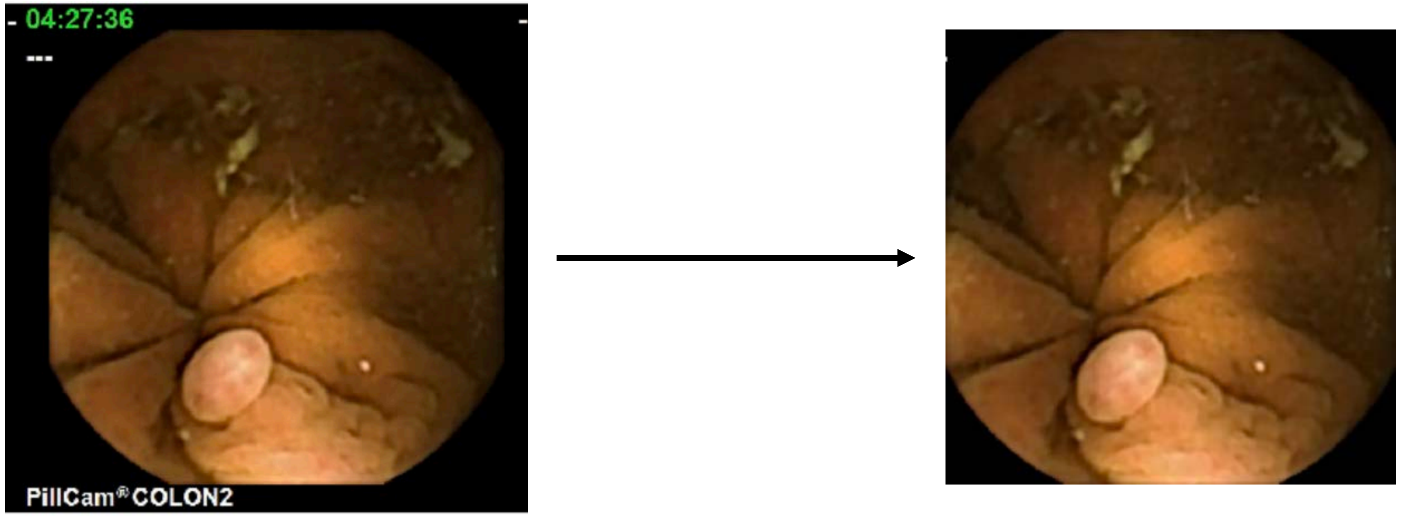
- The network pay attention to the corners, which may reduce the performance of the system.





DATABASE – PRE-PROCESSING

- Current pre-processing:



DATABASE – PRE-PROCESSING

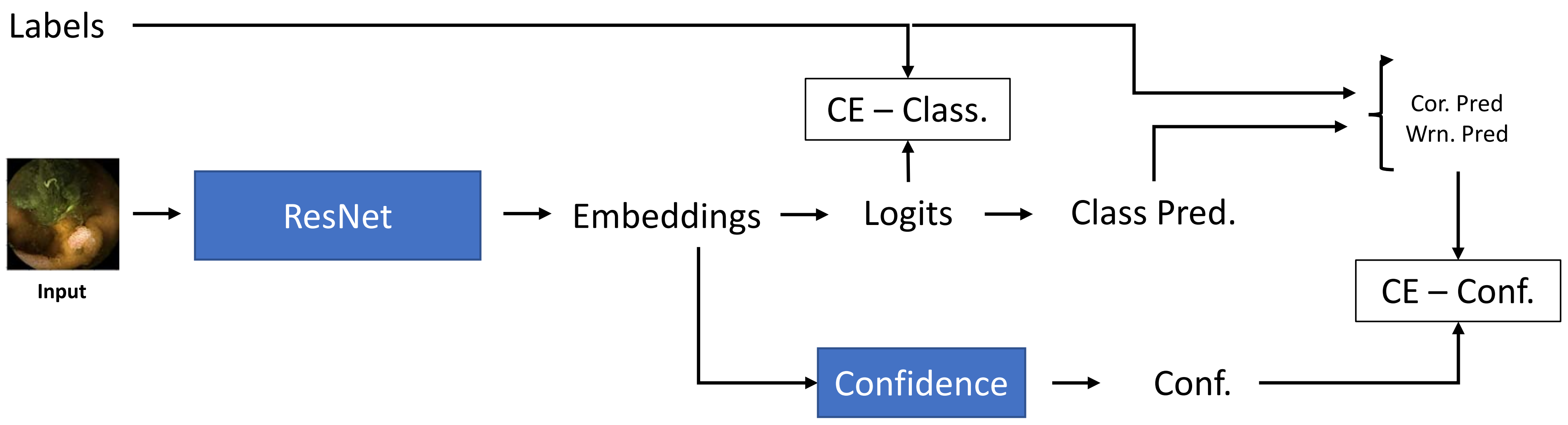
- About 20% of each image do not contain information (black regions).
- GOAL: Reduce black corners.[1]

[1] Xiaoqing Guo et al., 2020. **Semi-supervised WCE Image Classification with Adaptive Aggregated Attention**

What else ….

OVERESTIMATION OF THE NETWORK

- The predictions of the softmax tend to be overconfident.
  - A confidence value should help to increase the trustfulness in the system.



To be continued….