A new dataset of distortions on Wireless Capsule Endoscopy Images for pathological ifentification

Tan Sy NGUYEN

tansy.nguyen@math.univ-paris13.fr

LAGA, L2TI
Université Sorbonne Paris Nord

October 4, 2021

Overview

1. Objectives
   - Context
   - Wireless Capsule Endoscopy
     - Challenges
     - Solutions
2. Existing datasets
   - Existing GI datasets
   - HyperKvasir dataset
3. Our work
   - Method
   - Results

Objectives

The main objective of the project is to develop a smart system for:

- Identify the pathological finding on wireless capsule endoscopy (WCE) images
  - Including a pre-processing module that aims at improving the quality of the acquired images
  - Develop a set of image quality enhancement solutions based on kinds of distortion

There are **many kinds of distortion** & in **different levels**

Context

Alert

Colorectal cancer is a major health problem.

[1] Bray F, Ferlay J, Soerjomataram I, Siegel RL, Torre LA, Jemal A, "Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries", CA Cancer J Clin. 2018 Nov; 68(6):394-424.

[2] Santé Publique France, https://www.santepubliquefrance.fr/maladies-et-traumatismes/cancers/cancer-du-colon-rectum

[3] McKESSON, "Colorectal Cancer & Laboratory Screening", 2018

Context

Alert

Colorectal cancer is a major health problem.

Example

In 2018, the Colorectal cancer (CRC) is the third (second respectively) leading cause of cancer death in the world (France, respectively).[1,2]

1 Bray F, Ferlay J, Soerjomataram I, Siegel RL, Torre LA, Jemal A, "Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries", CA Cancer J Clin. 2018 Nov; 68(6):394-424.

2 Santé Publique France, https://www.santepubliquefrance.fr/maladies-et-traumatismes/cancers/cancer-du-colon-rectum

3 McKESSON, "Colorectal Cancer & Laboratory Screening", 2018

Context

Alert

Colorectal cancer is a major health problem.

Example

In 2018, the Colorectal cancer (CRC) is the third (second respectively) leading cause of cancer death in the world (France, respectively).[1,2]

Solution

Studies have shown that early detection can result in up to a **92% survival rate for stage I of cancer.**[3]

[1] Bray F, Ferlay J, Soerjomataram I, Siegel RL, Torre LA, Jemal A, "Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries", CA Cancer J Clin. 2018 Nov; 68(6):394-424.

[2] Santé Publique France, https://www.santepubliquefrance.fr/maladies-et-traumatismes/cancers/cancer-du-colon-rectum

[3] McKESSON, "Colorectal Cancer & Laboratory Screening", 2018

Wireless Capsule Endoscopy

Traditional endoscopy is often unpleasant and uncomfortable for the patient, can be painful, often requires moderate or deep sedation

Wireless Capsule Endoscopy

Traditional endoscopy is often unpleasant and uncomfortable for the patient, can be painful, often requires moderate or deep sedation

Wireless capsule endoscopy include its **non-invasive** character and its ability to visualize proximal and distal parts of the intestine

Challenges

- Some common acquisition distortions (**noise**, **blur**, **uneven illumination**, **specular reflection**) may affect the WCE based diagnosis.

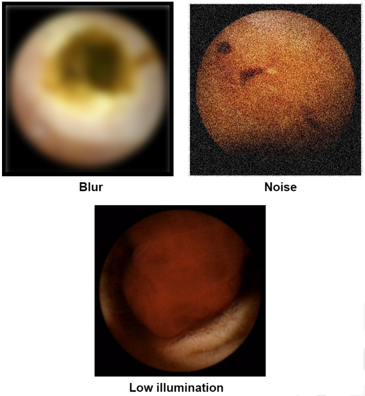


Figure 1: Illustration of some common WCE images distortions

Algorithm

input : *distorted_image*
output : *enhanced_image*

types_distortion = classifier (*distorted_image*);
for *type* **in** *types_distortion* **do**
    enhanced_image ← enhancer_{type} (*distorted_image*)
end
return *enhanced_image*

Algorithm

input : *distorted_image*
output : *enhanced_image*
types_distortion = $\boxed{\text{classifier}}$(*distorted_image*);
for *type* **in** *types_distortion* **do**
    enhanced_image ← $\boxed{\text{enhancer}_{type}}$(*distorted_image*)
end
return *enhanced_image*

Requirement:

Create the $\boxed{\text{classifier}}$ and $\boxed{\text{enhancer}_{type}}$

Algorithm

input : *distorted_image*
output : *enhanced_image*

$types\_distortion = \boxed{\text{classifier}}(distorted\_image)$;
for *type* **in** *types_distortion* **do**
    $enhanced\_image \leftarrow \boxed{\text{enhancer}_{type}}(distorted\_image)$
end
return *enhanced_image*

`Requirement:`

Create the $\boxed{\text{classifier}}$ and $\boxed{\text{enhancer}_{type}}$ by using **learning - based** method

Algorithm

input : *distorted_image*
output : *enhanced_image*

$types\_distortion = \boxed{\text{classifier}}(distorted\_image);$

for *type* **in** *types_distortion* **do**

  $enhanced\_image \leftarrow \boxed{\text{enhancer}_{type}}(distorted\_image)$

end

return *enhanced_image*

`Requirement:`

Create the $\boxed{\text{classifier}}$ and $\boxed{\text{enhancer}_{type}}$ by using **learning - based** method

Creating a dataset is the **most important** thing to do

Existing datasets

Table 1: An overview of existing GI datasets.

Dataset	Findings	Size
CVC-356 [1]	Polyps	356 images
CVC-ClinicDB (also named CVC-612) [2]	Polyps	612 images
CVC-VideoClinicDB (also named CVC-12k) [1]	Polyps	11954 images
CVC-ColonDB [1]	Polyps	380 images
Endoscopy Artifact detection 2019 [3]	Endoscopic Artifacts	5,138 images
ASU-Mayo polyp database [4]	Polyps	18,781 images
ETIS-Larib Polyp DB [5]	Polyps	196 images
KID [6]	Angiectasia, bleeding, inflammations, polyps	2371 images and 47 videos
GIANA 2017 [7]	Polyps & Angiodysplasia	3462 images and 38 videos
GIANA 2018 [8]	Polyps & Small bowel lesions	8262 images and 38 videos
GASTROLAB [9]	GI lesions	Some 100s of images andfew videos
WEO Clinical Endoscopy Atlas [10]	GI lesions	152 images
GI Lesions in Regular Colonoscopy Data Set [11]	GI lesions	76 images
Atlas of Gastrointestinal Endoscope [12]	GI lesions	1295 images
El salvador atlas of gastrointestinal video endoscopy [13]	GI lesions	5071 video clips
Kvasir [14]	Polyps, esophagitis, ulcerative colitis, Z-line,pylorus cecum, dyed polyp, dyed resection margins, stool	8000 images
Kvasir-SEG [15]	Polyps	1000 images
Nerthus [16]	Stool - categorization of bowel cleanliness	21 videos

Existing datasets

Table 1: An overview of existing GI datasets.

Dataset	Findings	Size
CVC-356 [1]	Polyps	356 images
CVC-ClinicDB (also named CVC-612) [2]	Polyps	612 images
CVC-VideoClinicDB (also named CVC-12k) [1]	Polyps	11954 images
CVC-ColonDB [1]	Polyps	380 images
Endoscopy Artifact detection 2019 [3]	Endoscopic Artifacts	5,138 images
ASU-Mayo polyp database [4]	Polyps	18,781 images
ETIS-Larib Polyp DB [5]	Polyps	196 images
KID [6]	Angiectasia, bleeding, inflammations, polyps	2371 images and 47 videos
GIANA 2017 [7]	Polyps & Angiodysplasia	3462 images and 38 videos
GIANA 2018 [8]	Polyps & Small bowel lesions	8262 images and 38 videos
GASTROLAB [9]	GI lesions	Some 100s of images andfew videos
WEO Clinical Endoscopy Atlas [10]	GI lesions	152 images
GI Lesions in Regular Colonoscopy Data Set [11]	GI lesions	76 images
Atlas of Gastrointestinal Endoscope [12]	GI lesions	1295 images
El salvador atlas of gastrointestinal video endoscopy [13]	GI lesions	5071 video clips
Kvasir [14]	Polyps, esophagitis, ulcerative colitis, Z-line,pylorus cecum, dyed polyp, dyed resection margins, stool	8000 images
Kvasir-SEG [15]	Polyps	1000 images
Nerthus [16]	Stool - categorization of bowel cleanliness	21 videos

They are rather small, and often limited to polyps. Several of these have also lately become unavailable.

Existing datasets

Table 1: An overview of existing GI datasets.

Dataset	Findings	Size
CVC-356 [1]	Polyps	356 images
CVC-ClinicDB (also named CVC-612) [2]	Polyps	612 images
CVC-VideoClinicDB (also named CVC-12k) [1]	Polyps	11954 images
CVC-ColonDB [1]	Polyps	380 images
Endoscopy Artifact detection 2019 [3]	Endoscopic Artifacts	5,138 images
ASU-Mayo polyp database [4]	Polyps	18,781 images
ETIS-Larib Polyp DB [5]	Polyps	196 images
KID [6]	Angiectasia, bleeding, inflammations, polyps	2371 images and 47 videos
GIANA 2017 [7]	Polyps & Angiodysplasia	3462 images and 38 videos
GIANA 2018 [8]	Polyps & Small bowel lesions	8262 images and 38 videos
GASTROLAB [9]	GI lesions	Some 100s of images andfew videos
WEO Clinical Endoscopy Atlas [10]	GI lesions	152 images
GI Lesions in Regular Colonoscopy Data Set [11]	GI lesions	76 images
Atlas of Gastrointestinal Endoscope [12]	GI lesions	1295 images
El salvador atlas of gastrointestinal video endoscopy [13]	GI lesions	5071 video clips
Kvasir [14]	Polyps, esophagitis, ulcerative colitis, Z-line,pylorus cecum, dyed polyp, dyed resection margins, stool	8000 images
Kvasir-SEG [15]	Polyps	1000 images
Nerthus [16]	Stool - categorization of bowel cleanliness	21 videos

They are rather small, and often limited to polyps. Several of these have also lately become unavailable.

Using **HyperKvasir** [17] dataset

HyperKvasir dataset

Table 2: Overview of the data records in the HyperKvasir dataset.

Data Record	# Files	Description
Labeled images	10,662 images	23 classes of findings
Segmented Images	1,000 images	Segmentation masks for polyp findings
Unlabeled Images	99,417 images	Unlabeled
Videos	374 videos	30 different classes

HyperKvasir dataset

Table 2: Overview of the data records in the HyperKvasir dataset.

Data Record	# Files	Description
Labeled images	10,662 images	23 classes of findings
Segmented Images	1,000 images	Segmentation masks for polyp findings
Unlabeled Images	99,417 images	Unlabeled
Videos	374 videos	30 different classes

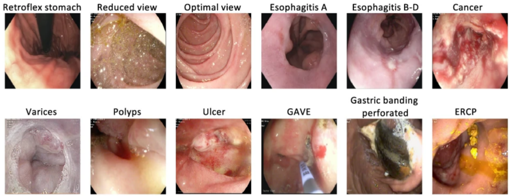
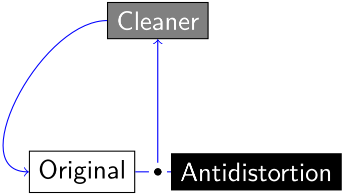


Figure 2: Image examples of the various labeled classes for images and/or videos.

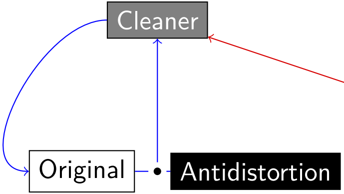
Ourwork



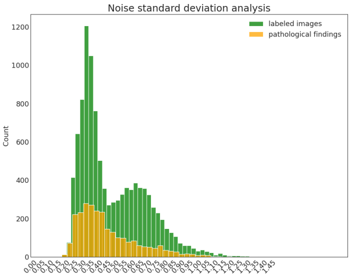
a) Clean the image

Step 1 Cleaning the existing distortion in HyperKvasir dataset

Ourwork



a) Clean the image



Step 1 Cleaning the existing distortion in HyperKvasir dataset

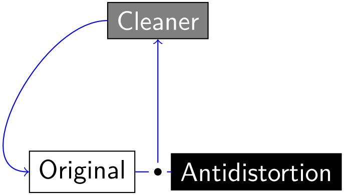
Ourwork

a) Clean the image

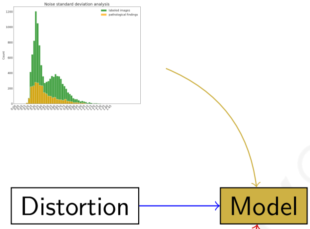
b) Create model

Step 2 Creating the model to generate the new artificial distortions

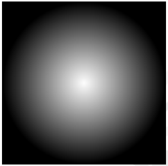
Ourwork



a) Clean the image



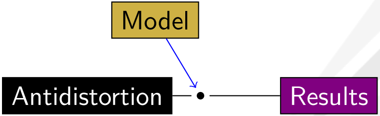
b) Create model



Ourwork

Step 3 Add the new artificial distortions to the antidistorted images

a) Clean the image *b) Create model*



c) Add artificial distortion

Results

Stage 1: we have to clean the existing distortion in the HyperKvasir dataset.

Results

Stage 1: we have to clean the existing distortion in the HyperKvasir dataset.

Original Image → Denoising

Results

Stage 1: we have to clean the existing distortion in the HyperKvasir dataset.

Original Image → **Denoising** → Debluring

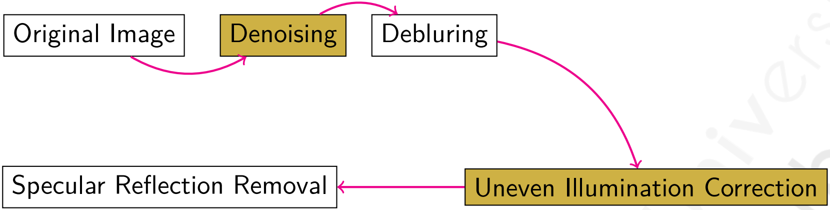
Results

Stage 1: we have to clean the existing distortion in the HyperKvasir dataset.

Original Image → Denoising → Debluring → Uneven Illumination Correction

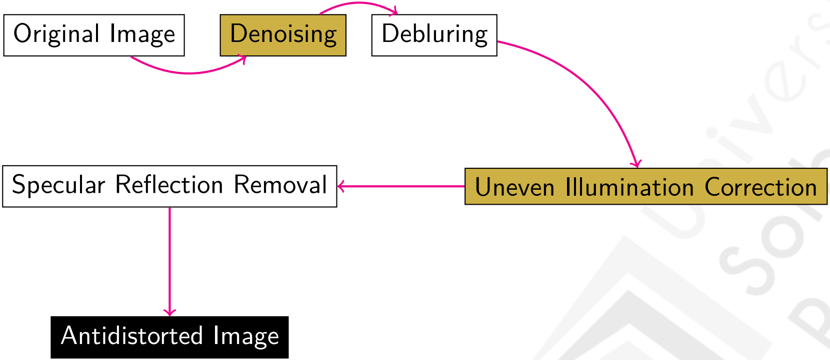
Results

Stage 1: we have to clean the existing distortion in the HyperKvasir dataset.



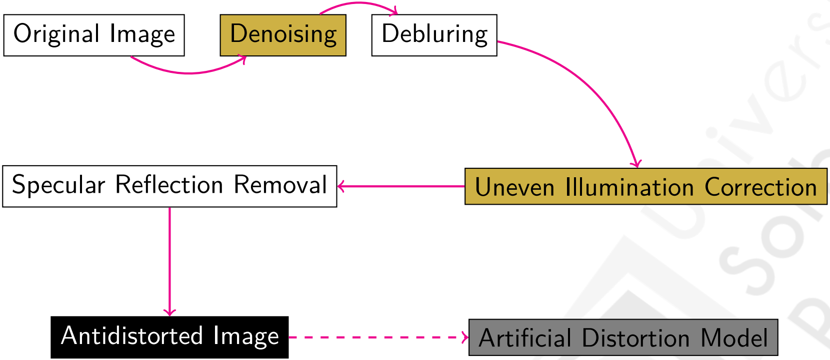
Results

Stage 1: we have to clean the existing distortion in the HyperKvasir dataset.



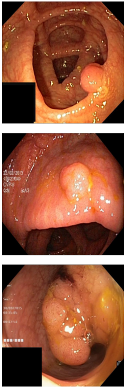
Results

Stage 1: we have to clean the existing distortion in the HyperKvasir dataset.



Results

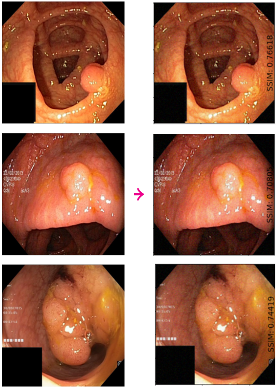
- Noise



a) Original

Results

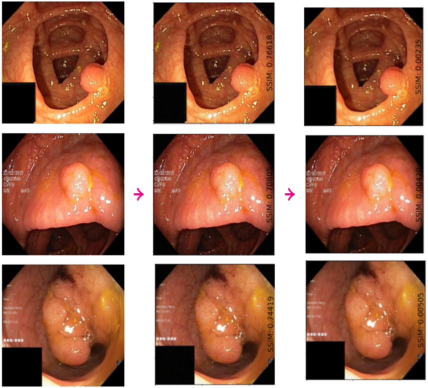
- Noise



a) Original *b) Noisy*

Results

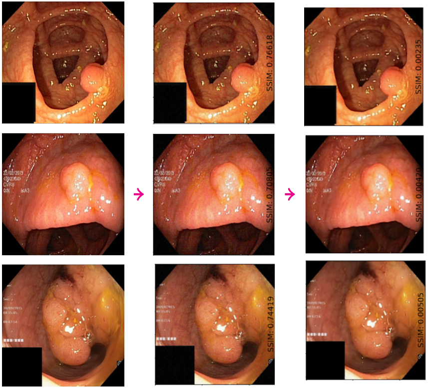
- Noise



a) Original *b) Noisy* *c) Denoised*

Results

- Noise



a) Original *b) Noisy* *c) Denoised*

		Noisy-image	C-BM3D	NM-Mean	CAE
σ_n	0.25	0.8095	0.7364	0.9068	**0.9158**
	0.3	0.7561	0.7275	0.8884	**0.9097**
	0.35	0.7084	0.7106	0.8752	**0.9033**
	0.4	0.6615	0.7039	0.8614	**0.8991**
	0.6	0.5102	0.6517	0.8222	**0.8766**
	0.65	0.4810	0.6428	0.8132	**0.8681**
	0.7	0.4508	0.6389	0.8051	**0.8661**

d) Comparison using mean SSIM for different level where $n \sim N(0, \sigma_n^2)$

Results

- Noise

Table 3: Comparison using mean SSIM for different level where $n \sim N(0, \sigma_n^2)$

	σ_n						
	0.25	0.3	0.35	0.4	0.6	0.65	0.7
Noisy-image	0.8095	0.7561	0.7084	0.6615	0.5102	0.4810	0.4508
C-BM3D	0.7364	0.7275	0.7106	0.7039	0.6517	0.6428	0.6389
NM-Mean	0.9068	0.8884	0.8752	0.8614	0.8222	0.8132	0.8051
CAE	**0.9158**	**0.9097**	**0.9033**	**0.8991**	**0.8766**	**0.8681**	**0.8661**

Table 4: Comparison using mean PSNR for different level where $n \sim N(0, \sigma_n^2)$

	σ_n						
	0.25	0.3	0.35	0.4	0.6	0.65	0.7
Noisy-image	32.89	31.44	30.34	29.32	26.37	25.85	25.22
C-BM3D	27.51	27.52	27.47	27.47	27.28	27.21	27.16
NM-Mean	**35.81**	**34.74**	**34.12**	**33.48**	**31.85**	**31.51**	31.14
CAE	32.28	32.56	32.70	32.38	31.42	31.18	**31.43**

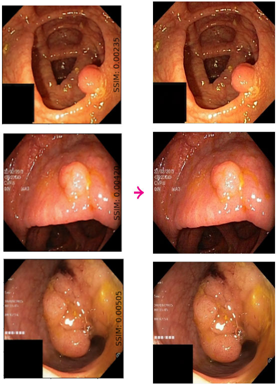
Results

- Blur

a) Denoised

Results

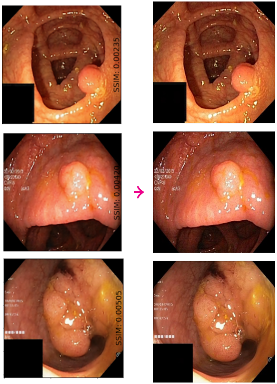
- Blur



a) Denoised *b) Deblurred*

Results

- Blur



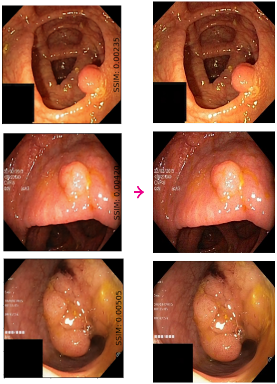
a) Denoised *b) Deblurred*

$$index = var(\mathcal{L}(f(x, y)))$$

*Apply the **variance of the Laplacian**[18] method to your own photos to detect the amount of blurring.*

Results

- Blur



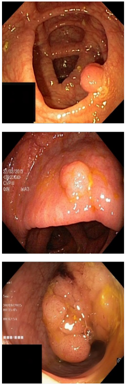
a) Denoised *b) Deblurred*

	Denoised-image	Deblured- mage
First exp	378	501
Second exp	321	428
Third exp	224	367

Table 5: The variance Laplacian Index before and after deblurring.

Results

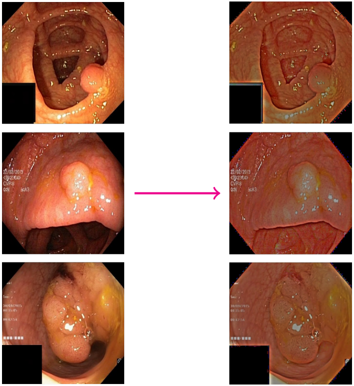
- Uneven Illumination



a) Deblurred

Results

- Uneven Illumination

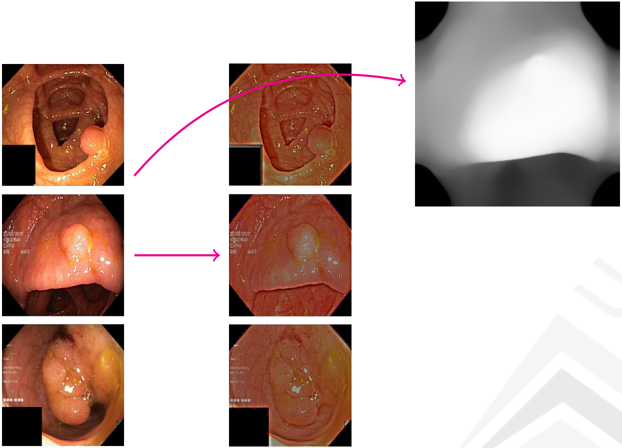


a) Deblurred

b) Uneven Illumination Correction

Results

- Uneven Illumination

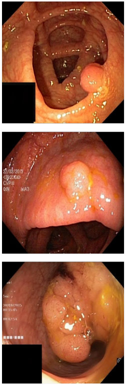


a) Deblurred

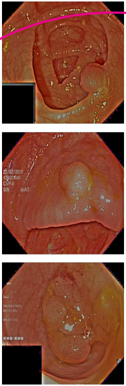
b) Uneven Illumination Correction

Results

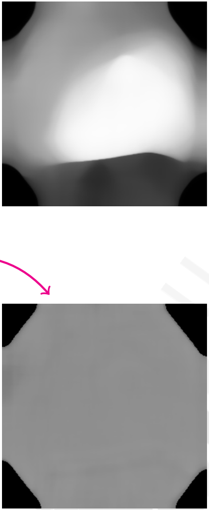
- Uneven Illumination



a) Deblurred

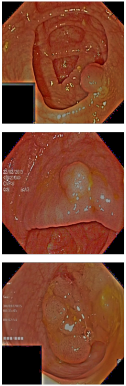


b) Uneven Illumination Correction



Results

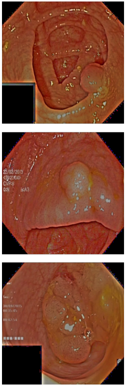
- Specular Reflection



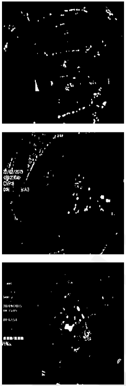
a) Uneven Illumination Correction

Results

- Specular Reflection

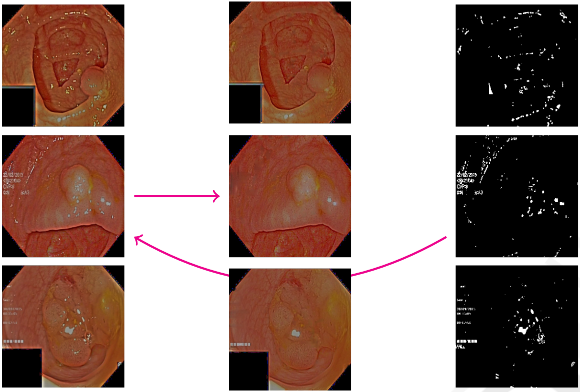


a) Uneven Illumination Correction



Results

- Specular Reflection



a) Uneven Illumination Correction

b) Specular Reflection Inpainting

Results

Stage 2: We add the artificial distortion to the image

Antidistorted Image

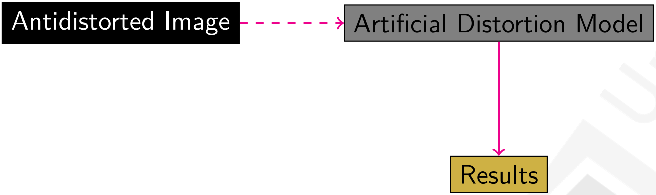
Results

Stage 2: We add the artificial distortion to the image

Antidistorted Image ⇢ Artificial Distortion Model

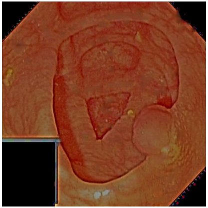
Results

Stage 2: We add the artificial distortion to the image



Results

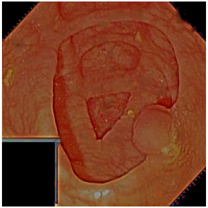
- Noise



a) Antidistorted Image

Results

- Noise

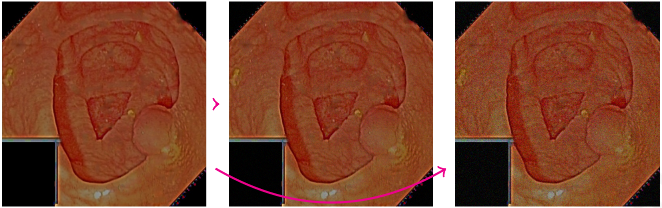


a) Antidistorted Image

b) Noised image with Gaussian Noise $n \sim N(0, \sigma_n^2 = (0.0005)^2)$

Results

- Noise



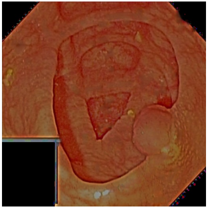
a) Antidistorted Image

b) Noised image with Gaussian Noise $n \sim N(0, \sigma_n^2 = (0.0005)^2)$

c) Noised image with Gaussian Noise $n \sim N(0, \sigma_n^2 = (0.005)^2)$

Results

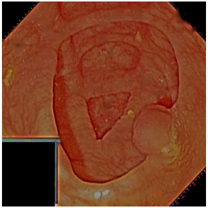
- Blur

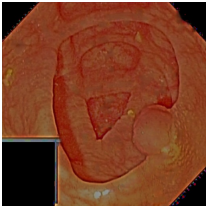


a) Antidistorted Image

Results

- Blur



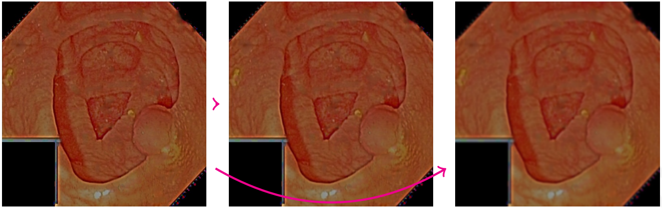


a) Antidistorted Image

b) Blured image with Defocus Blur $b \sim N(0, \sigma_b^2 = (0.75)^2)$

Results

- Blur



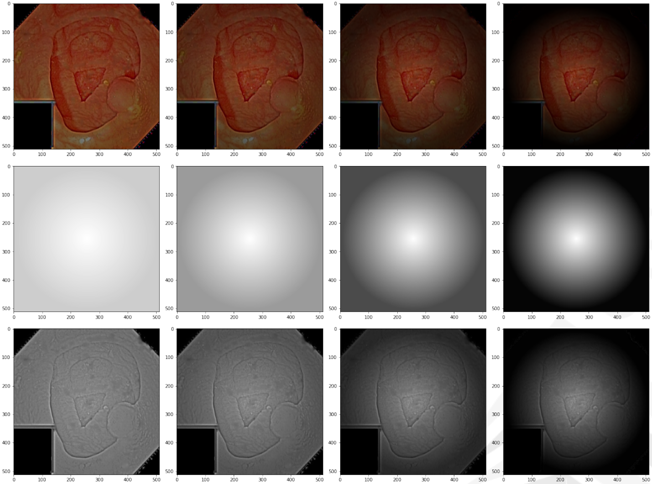
a) Antidistorted Image

b) Blured image with Defocus Blur $b \sim N(0, \sigma_b^2 = (0.75)^2)$

c) Blured image with Defocus Blur $b \sim N(0, \sigma_b^2 = (2)^2)$

Results

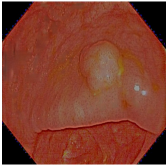
- Uneven Illumination



Artificial Uneven Illumination process

Results

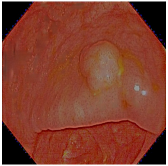
- Specular Reflection

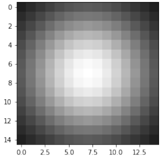


a) Antidistorted Images

Results

- Specular Reflection





a) Antidistorted Images

Results

- Specular Reflection

a) Antidistorted Images

b) Artificial Specular Reflection Image

References

[1] Jorge Bernal, F. J. Sánchez, and F. Vilariño. "Towards automatic polyp detection with a polyp appearance model". In: **Pattern Recognit.** 45 (2012), pp. 3166–3182.

[2] Jorge Bernal et al. "WM-DOVA maps for accurate polyp highlighting in colonoscopy: Validation vs. saliency maps from physicians". In: **Computerized medical imaging and graphics : the official journal of the Computerized Medical Imaging Society** 43 (2015), pp. 99–111.

[3] Sharib Ali et al. "Endoscopy artifact detection (EAD 2019) challenge dataset". In: **ArXiv** abs/1905.03209 (2019).

[4] Nima Tajbakhsh, S. Gurudu, and Jianming Liang. "Automated Polyp Detection in Colonoscopy Videos Using Shape and Context Information". In: **IEEE Transactions on Medical Imaging** 35 (2016), pp. 630–644.

[5] Juan Silva et al. "Toward embedded detection of polyps in WCE images for early diagnosis of colorectal cancer". In: **International Journal of Computer Assisted Radiology and Surgery** 9 (2013), pp. 283–293.

[6] A. Koulaouzidis et al. "KID Project: an internet-based digital video atlas of capsule endoscopy for research purposes". In: **Endoscopy International Open** 5 (2017), E477–E483.

[7] Y. Guo and B. Matuszewski. "GIANA Polyp Segmentation with Fully Convolutional Dilation Neural Networks". In: **VISIGRAPP** (2019).

[8] Jorge Bernal et al. "Polyp Detection Benchmark in Colonoscopy Videos using GTCreator: A Novel Fully Configurable Tool for Easy and Fast Annotation of Image Databases". In: (2018).

[9] the gastrointestinal site. "Gastrolab". In: **http://www.gastrolab.net/index.htm. Accessed: 2019-12-12** (2019).

[10] Endoatlas. "Weo clinical endoscopy atlas". In: **http://www.endoatlas.org/index.php. Accessed: 2019-12-12** (2019).

[11] UCI. "Gastrointestinal lesions in regular colonoscopy dataset". In: (2019). URL: `http://www.depeca.uah.es/colonoscopy_dataset/`.

[12] Atlas. "The atlas of gastrointestinal endoscope". In: (2019). URL: `http://www.endoatlas.com/atlas_1.html`.

[13] Atlas. " El salvador atlas of gastrointestinal video endoscopy". In: (2019). URL: `http://www.gastrointestinalatlas.com/index.html`.

[14] Konstantin Pogorelov et al. "KVASIR: A Multi-Class Image Dataset for Computer Aided Gastrointestinal Disease Detection". In: **Proceedings of the 8th ACM on Multimedia Systems Conference** (2017).

[15] Debesh Jha et al. "Kvasir-SEG: A Segmented Polyp Dataset". In: **ArXiv** abs/1911.07069 (2020).

[16] Konstantin Pogorelov et al. "Nerthus: A Bowel Preparation Quality Video Dataset". In: **Proceedings of the 8th ACM on Multimedia Systems Conference** (2017).

[17] Hanna Borgli et al. "HyperKvasir, a comprehensive multi-class image and video dataset for gastrointestinal endoscopy". In: **Scientific Data** 7 (2020).

[18] Said Pertuz, Domenec Puig, and Miguel Ángel García. "Analysis of focus measure operators for shape-from-focus". In: **Pattern Recognit.** 46 (2013), pp. 1415–1432.

Thank you for watching!

Contact:

tansy.nguyen@math.univ-paris13.fr