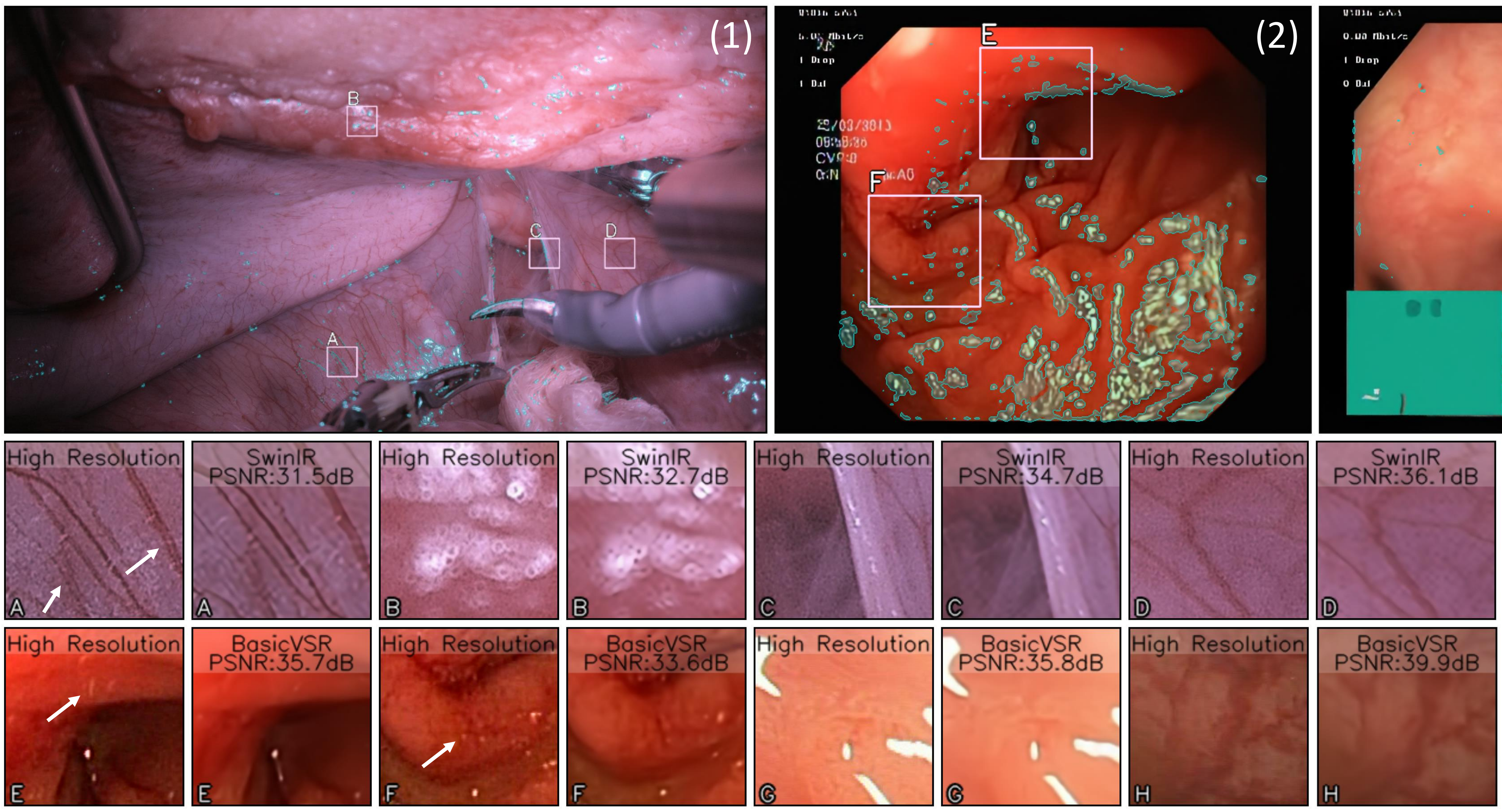
Trustworthy Endoscopic Super-Resolution

Julio Silva-Rodríguez · Ender Konukoglu - ETH Zürich

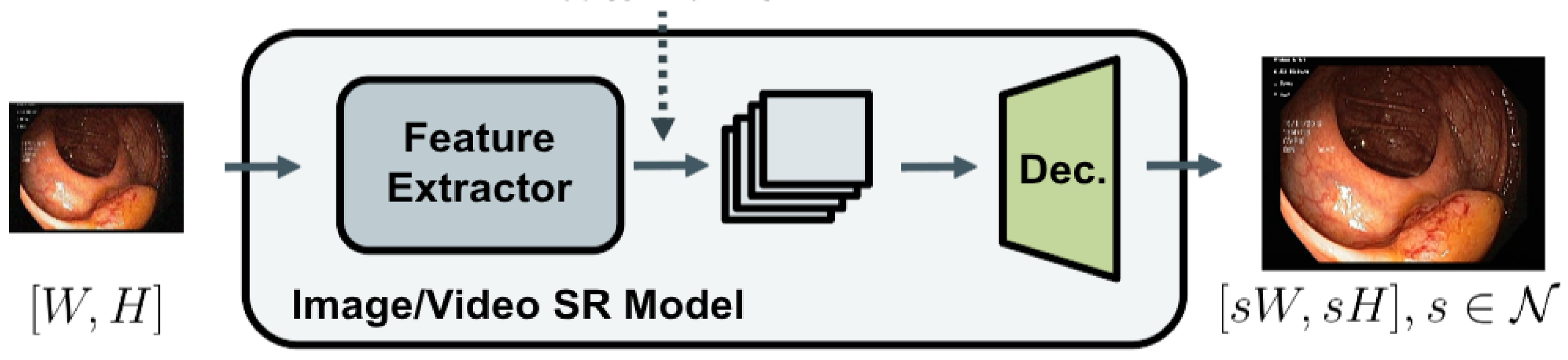
- **AI-based enhancement** techniques can improve image quality under real-time hardware constraints. However, these might produce **hallucinated outputs**, causing **reliability concerns**.
- We introduce **Conformal Failure Masks** (CFMs), which provide **operational failure decisions** with theoretical guarantees **controlled by the user**.



Qualitative visualizations of CFMs (blue). Masks are created at an error level of **22dB** and tolerating a **false-negative rate of 5%**.
(1): SurgiSR4K, using SwinIR-4x: 950x540p to 3840x2160p. **(2-3): HyperKvasir**, using BasicVSR-4x:180x144p to 750x576p.

Super-resolution for endoscopic imaging

- **Real-time image/video super-resolution (SR)** models are composed by a feature extractor and a decoder increasing resolution and denoising.



- Ideally, **well-established open-access models**, e.g. SwinIR or BasicVSR, could be directly used, avoiding retraining on large data collections.
- How a **practical and reliable SR system** should be?
  1. Image/video SR model-agnostic without retraining.
  2. Real-time latency.
  3. Operational failure decisions (in contrast to interval-valued outputs).
  4. Control and guarantees over detected failures.

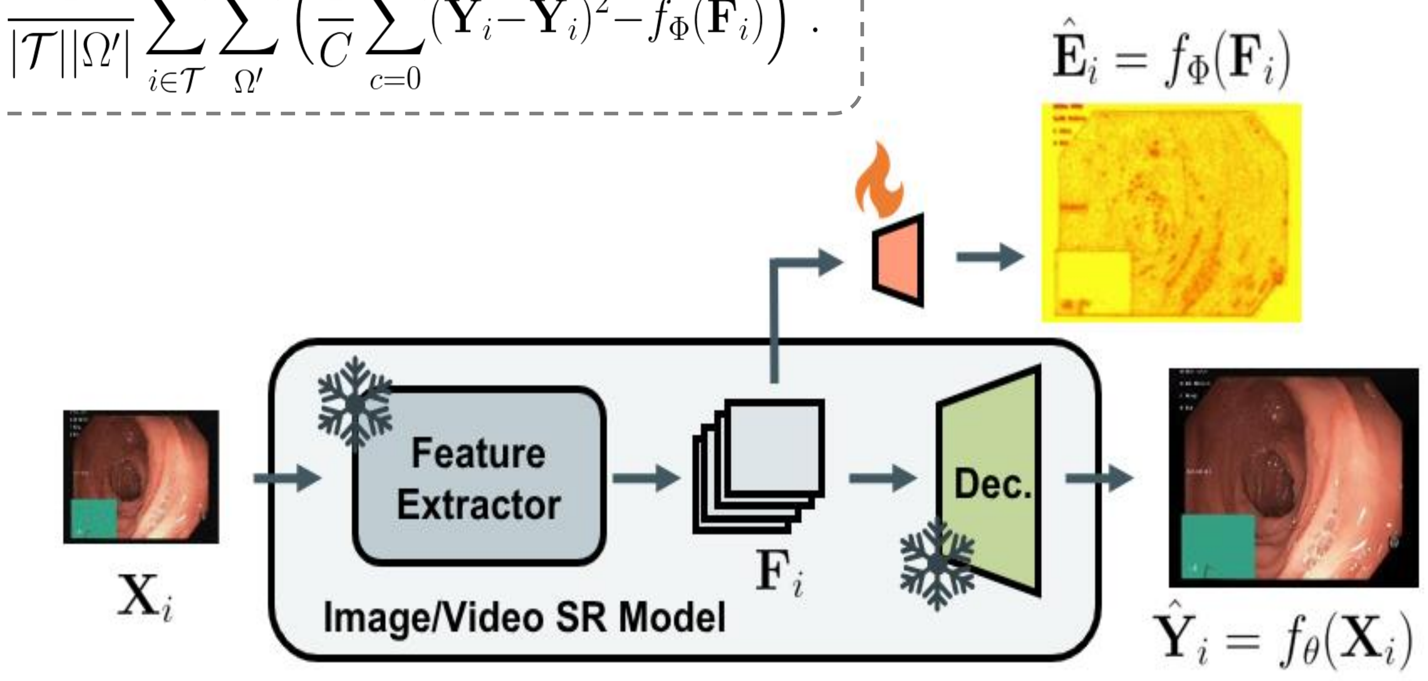
Existing UQ frameworks** [MC-Dropout, Ensembles, Bayesian, CQR]* ***cannot be deployed under these constraints.

	SR Method	$\alpha = 0.10$ Avg. PSNR	FNR (%)	Avg. Mask Size (%)	$\alpha = 0.05$ Avg. PSNR	FNR (%)	Avg. Mask size (%)
Surgi-2×	Bicubic	36.5 ± 1.2	–	–	36.5 ± 1.2	–	–
	SwinIR-L-2×	38.7 ± 1.5	–	–	38.7 ± 1.5	–	–
	+CFM ($\tau^{\text{fail}} = 22$ dB)	40.9 ± 0.8	10.8	2.7 ± 1.3	41.4 ± 0.7	5.3	4.0 ± 1.7
	+CFM ($\tau^{\text{fail}} = 24$ dB)	41.3 ± 0.7	10.5	3.6 ± 1.6	41.8 ± 0.7	5.1	5.4 ± 2.1
	+CFM ($\tau^{\text{fail}} = 26$ dB)	41.7 ± 0.7	10.4	4.8 ± 2.0	42.3 ± 0.7	5.0	7.4 ± 3.1
Surgi-4×	Bicubic	33.0 ± 1.8	–	–	33.0 ± 1.8	–	–
	SwinIR-L-4×	33.7 ± 1.9	–	–	33.7 ± 1.9	–	–
	+CFM ($\tau^{\text{fail}} = 22$ dB)	34.6 ± 1.2	9.6	4.0 ± 4.2	34.7 ± 1.2	4.6	5.4 ± 5.2
	+CFM ($\tau^{\text{fail}} = 24$ dB)	34.7 ± 1.2	9.5	5.0 ± 4.9	34.8 ± 1.2	4.5	6.9 ± 6.0
	+CFM ($\tau^{\text{fail}} = 26$ dB)	34.9 ± 1.2	9.3	8.9 ± 6.6	35.1 ± 1.2	5.1	17.7 ± 9.7
HKvasir	Bicubic	26.0 ± 0.7	–	–	26.0 ± 0.7	–	–
	BasicVSR	30.4 ± 1.2	–	–	30.4 ± 1.2	–	–
	+CFM ($\tau^{\text{fail}} = 22$ dB)	37.4 ± 2.1	10.2	5.6 ± 2.6	38.3 ± 2.1	5.0	10.0 ± 4.3
	+CFM ($\tau^{\text{fail}} = 24$ dB)	37.9 ± 2.1	10.2	7.5 ± 3.4	38.7 ± 2.0	5.0	13.7 ± 5.3
	+CFM ($\tau^{\text{fail}} = 26$ dB)	38.3 ± 2.1	10.2	10.3 ± 4.4	39.2 ± 2.0	5.0	19.4 ± 6.3

Step 1: Lightweight SR error quantification

- A **low-latency network** based on convolutional blocks is trained to **estimate the reconstruction error**, using **frozen intermediate SR model features**, and few LR-HR paired videos.

$$\min_{\Phi} \frac{1}{|\mathcal{T}||\Omega'|} \sum_{i \in \mathcal{T}} \sum_{\Omega'} \left(\frac{1}{C} \sum_{c=0}^{C-1} (\hat{\mathbf{Y}}_i - \mathbf{Y}_i)^2 - f_{\Phi}(\mathbf{F}_i) \right)^2.$$



Step 2: Conformal Failure Masks

- An **operational threshold** transforms the estimate into **binary masks**.

$$\hat{\mathbf{M}}_i^{\tau} = \mathbf{1}\{\hat{\mathbf{E}}_i \geq \tau\} \in \{0,1\}^{\Omega' \times 1}$$

- We **control**: the desired **error level** (PSNR) that the practitioner requires and the **false-negative rate** tolerance on detected failures (FNR):

$$\mathrm{FNR}(\tau^{\text{fail}}, \tilde{\tau}) \leq \alpha.$$

$$\mathrm{FNR}_{\mathcal{D}_{\text{cal}}}(\tau^{\text{fail}}, \tau) = \frac{1}{N} \sum_{i \in \mathcal{C}} \sum_{\Omega'} \mathbf{1}\{\mathbf{M}_i^{\tau^{\text{fail}}} = 1 \wedge \hat{\mathbf{M}}_i^{\tau} = 0\}$$

- Using **fresh calibration data** and **conformal risk control** principles, we provide **theoretical guarantees** upon exchangeability of incoming data.

$$\tilde{\tau} = \sup \left\{ \tau : \frac{N}{N+1} \mathrm{FNR}_{\mathcal{D}_{\text{cal}}}(\tau^{\text{fail}}, \tau) - \frac{1}{N+1} \leq \alpha \right\}.$$

Acknowledgements. Innovation project supported by Innosuisse (122.308 IP-LS).