OfGAN: Realistic Rendition of Synthetic Colonoscopy Videos

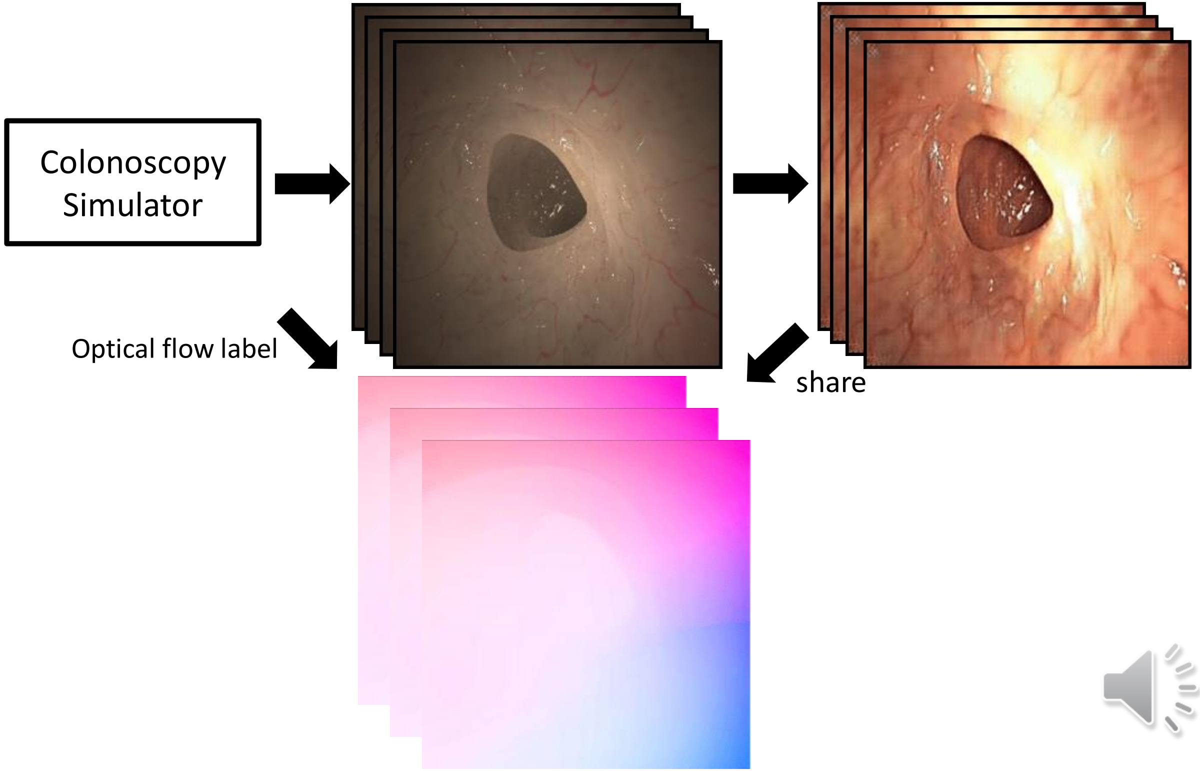
Jiabo Xu [1, 2], Saeed Anwar [1, 2], Nick Barnes [1], Florian Grimpen [3], Olivier Salvado [1] , Stuart Anderson [1], Mohammad Ali Armin [1, 2]

[1] Data61, Commonwealth Scientific Industrial Research Organization of Australia (CSIRO)
[2] Research School of Engineering, Australian National University
[3] Department of Gastroenterology and HRBW Hospital, Brisbane, Australia

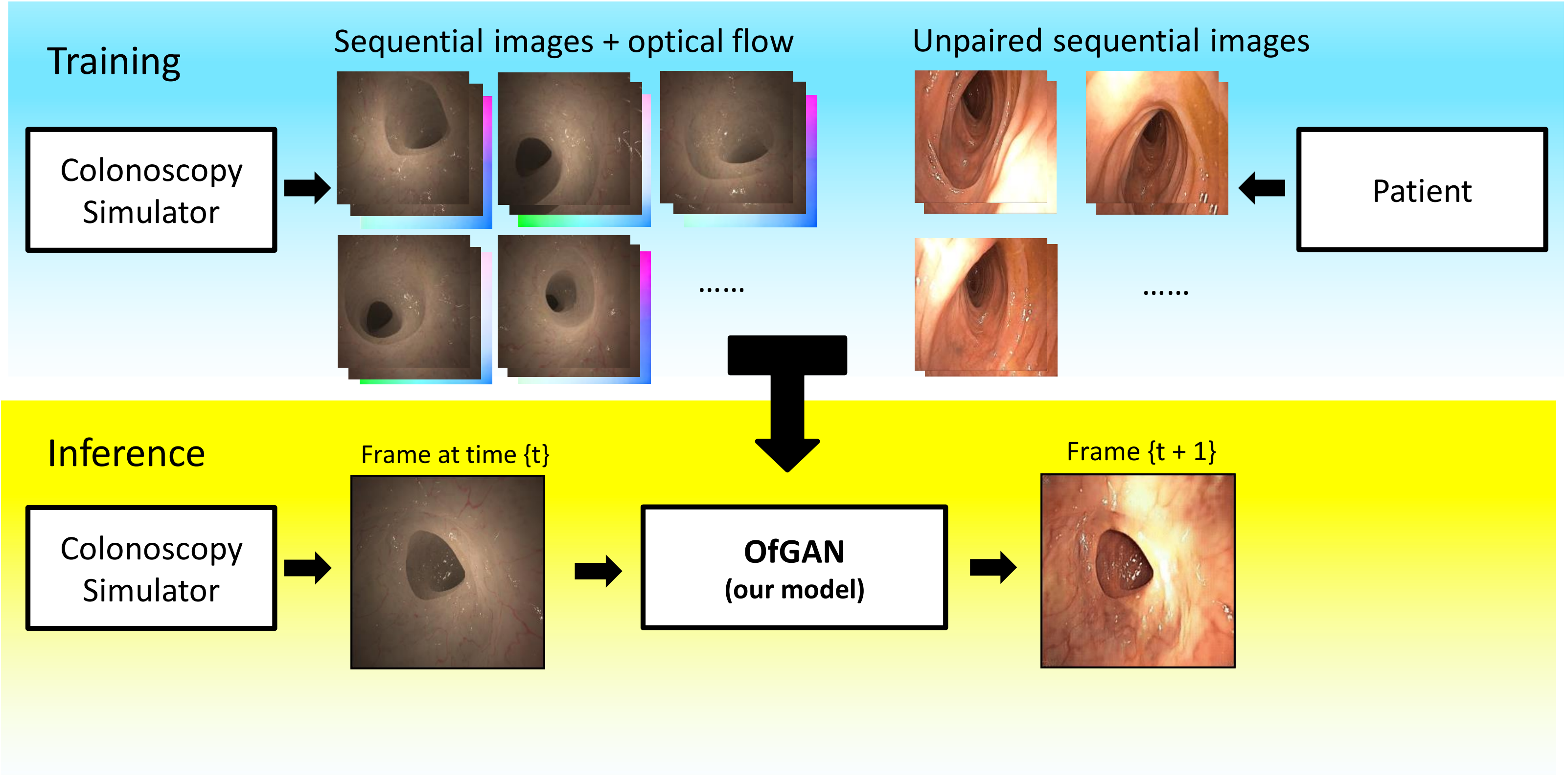
Australia's National Science Agency

Motivation

- Data insufficiency
  - Hard to label
  - Patient-specific
  - Privacy policy
- Simulation technology
  - Auto-rendering



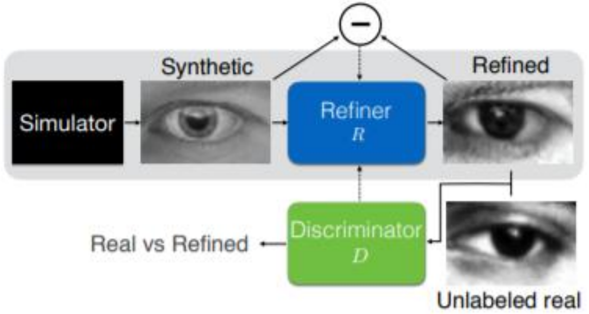
Task



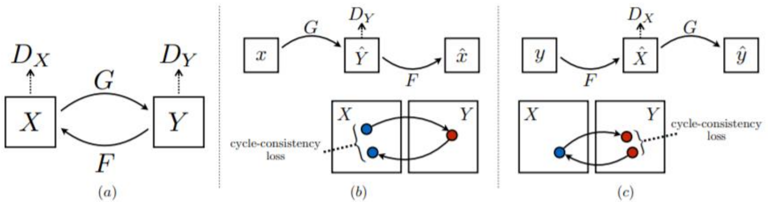
Training
Sequential images + optical flow
Unpaired sequential images
Colonoscopy Simulator
Patient
......
......
Inference
Frame at time {t}
Frame {t + 1}
Colonoscopy Simulator
OfGAN
(our model)

Related works & Challenges

Domain Adaptation: S+U GAN
Unpaired Image-to-Image Translation: CycleGAN



S+U GAN, Shrivastava et al. 2017



CycleGAN, Zhu et al. 2017

Challenges:

- Domain transformation from synthetic to real
- Keep temporal-consistent simultaneously
- Preserve the structure of the input

Methodology – overall

GAN-based model

- enable domain transformation

Cycle-consistent structure (CycleGan)

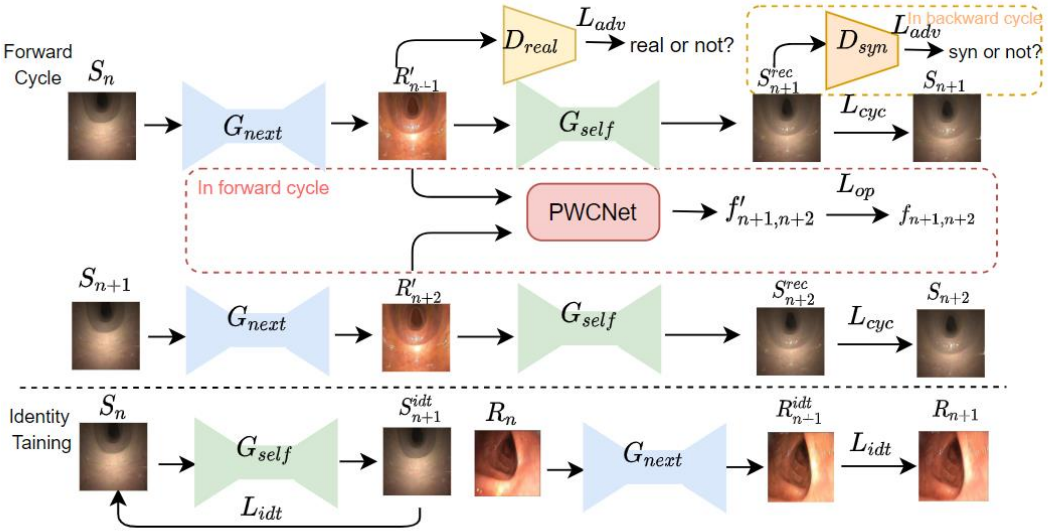
- constrain target distribution

Optical flow loss

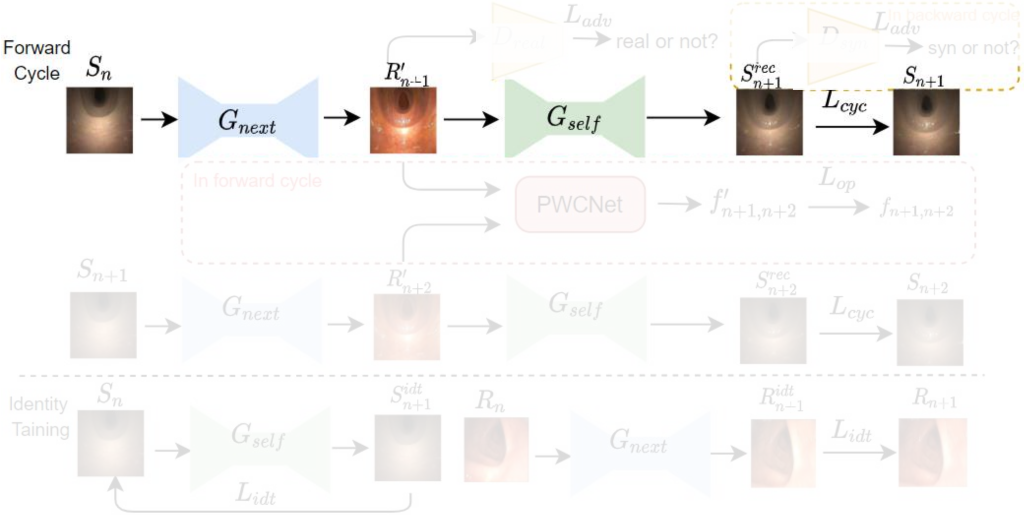
- generalize temporal information

Temporal consistent structure

- build up connections



Methodology – temporal consistent structure



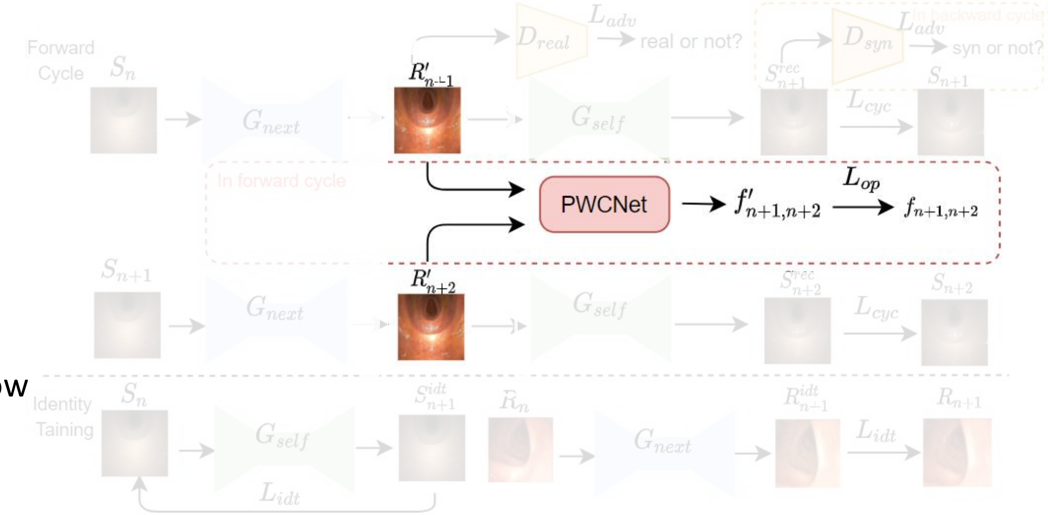
Novel temporal mapping chain:

$$S_n \xrightarrow{G_{next}} R'_{n+1} \xrightarrow{G_{self}} S^{rec}_{n+1}$$

Temporal consistent loss:

$$\begin{aligned}\mathcal{L}_{cyc}(G_{next}, G_{self}) =& \mathrm{E}_{s\sim Pdata(S)}[||G_{self}(G_{next}(s_n)) - s_{n+1}||_1]+\\ & \mathrm{E}_{r\sim Pdata(R)}[||G_{next}(G_{self}(r_m)) - r_{m+1}||_1].\end{aligned}$$

Methodology – optical flow loss

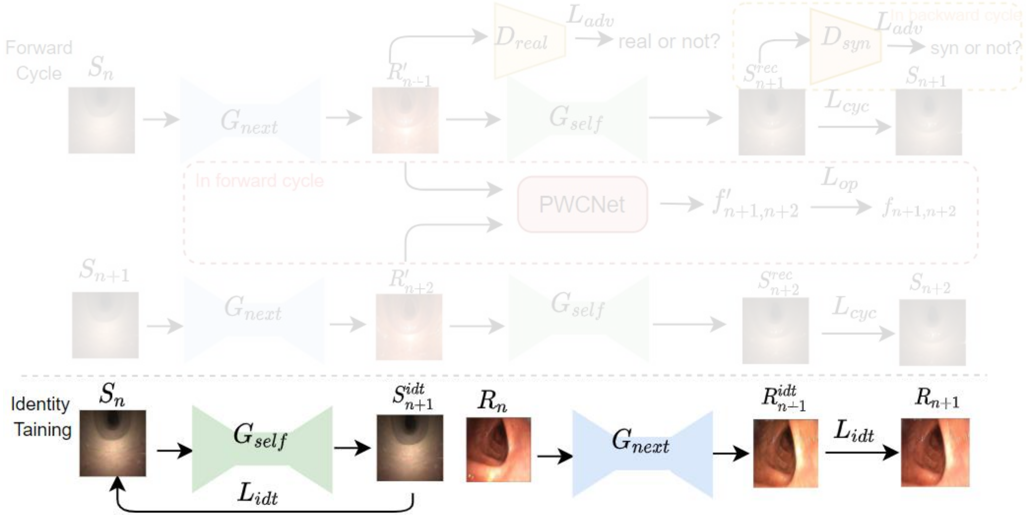


Used PWCNet (Sun et al., 2018) as the optical flow estimator $Op(\cdot)$

Optical flow loss:

$$\mathcal{L}_{op}(G_{next}) = \mathrm{E}_{s\sim Pdata(S), f\sim Pdata(F)}[||Op(r'_n, r'_{n+1}) - f_{n,n+1}||_1]$$

Methodology – perceptual identity loss



Use the output of the second conv-block of pretrained VGG as the feature extractor $\theta(\cdot)$.

Perceptual identity loss:

$$\begin{aligned}\mathcal{L}_{idt}(G_{next}, G_{self}) = & \mathrm{E}_{r\sim Pdata(R)}[\theta(G_{next}(r_m)), \theta(r_{m+1})] + \\ & \mathrm{E}_{s\sim Pdata(S)}[(\theta(G_{self}(s_n)), \theta(s_n)]\end{aligned}$$

Methodology – overall loss

Overall loss:

$$\mathcal{L}(G_{next}, G_{self}, D_{syn}, D_{real}) = \mathcal{L}_{adv}(G_{next}, G_{self}) + \lambda \mathcal{L}_{cyc}(G_{next}, G_{self}) + \beta \mathcal{L}_{idt}(G_{next}, G_{self}) + \underline{\sigma} \mathcal{L}_{op}(G_{next}),$$

Optimization on min-max problem of:

$$G^*_{next}, G^*_{self} = \arg \min_{G_{next}, G_{self}} \max_{D_{syn}, D_{real}} \mathcal{L}(G_{next}, G_{self}, D_{syn}, D_{real})$$

Experiment – datasets

State of the art CSIRO simulator:

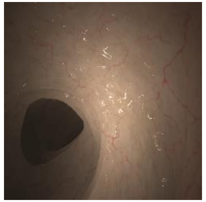
Train: Synthetic videos - 8000 video frames with ground truth optical flow from 5 different simulated colons

Test: 2000 frames from two unknown synthetic colonoscopy videos.

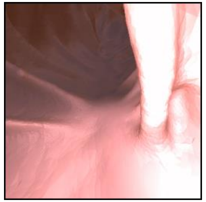
A Published CT colonoscopy dataset [23] (no GT optical flow)

Real video:

Consists of 1472 video frames (after cleaning) captured from patients by our specialists.



Synthetic



CT

Real

Experiment – ablation study

Ablation Study:

Model structure

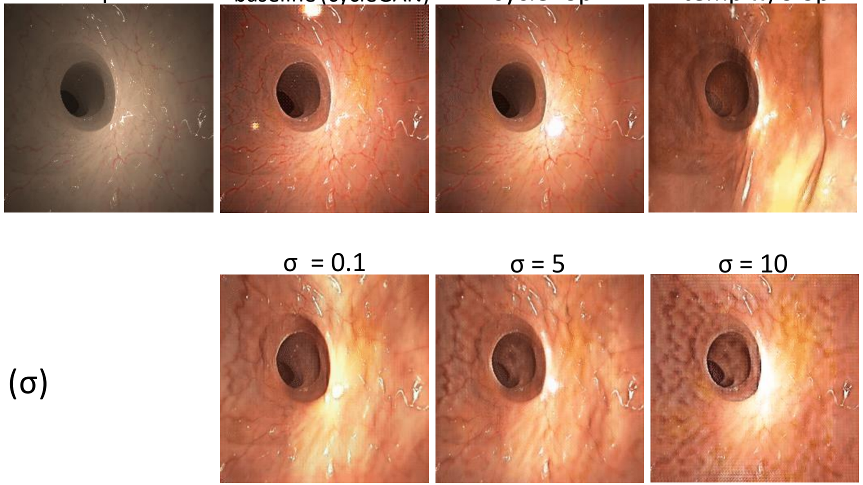
- Cycle (baseline)
- Cycle + op
- Temp w/o op
- Temp + op (proposed)

σ = 0.1 or 5 or 10

(weight of $\mathcal{L}_{op}$)

β = 75

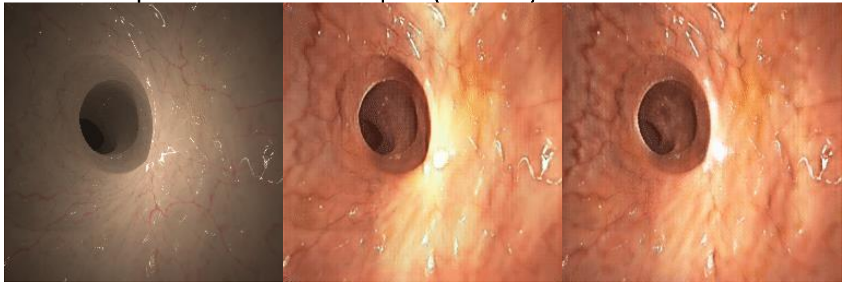
(weight of $\mathcal{L}_{idt}$)



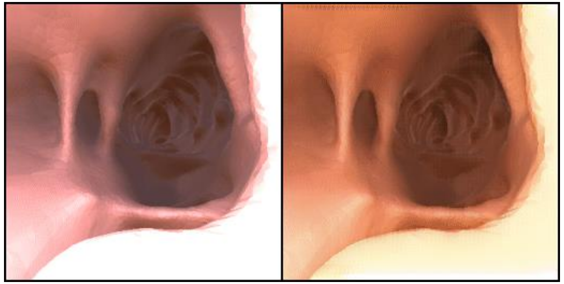
Experiment – our results

Input | output (σ = 0.1) | output (σ = 5)

Synthetic to real



CT to real
(no GT optical flow)



Experiment – features

- **Better single image quality**: Spatial + Temporal > Spatial
- **Generalizable**: one simulator only requires one trained model.
- **Fast inference**: "simulator + our model" generates realistic data without extra waiting.
- **Robust**: rarely need manual cleaning on generated dataset.

More examples on model with σ = 0.1

Conclusion

Summary:

- OfGAN successfully achieves high spatial and temporal quality on video domain transformation.

- The transformed data can be patient-specific.

- It can be easily applied on a given simulator for generating a large number of realistic data.

Limitations:

- There are small white spots on transformed frames.
- It is inevitable to bring on estimation error when using an optical flow estimator.
- We will verify our model on data from other medical simulators in the future.