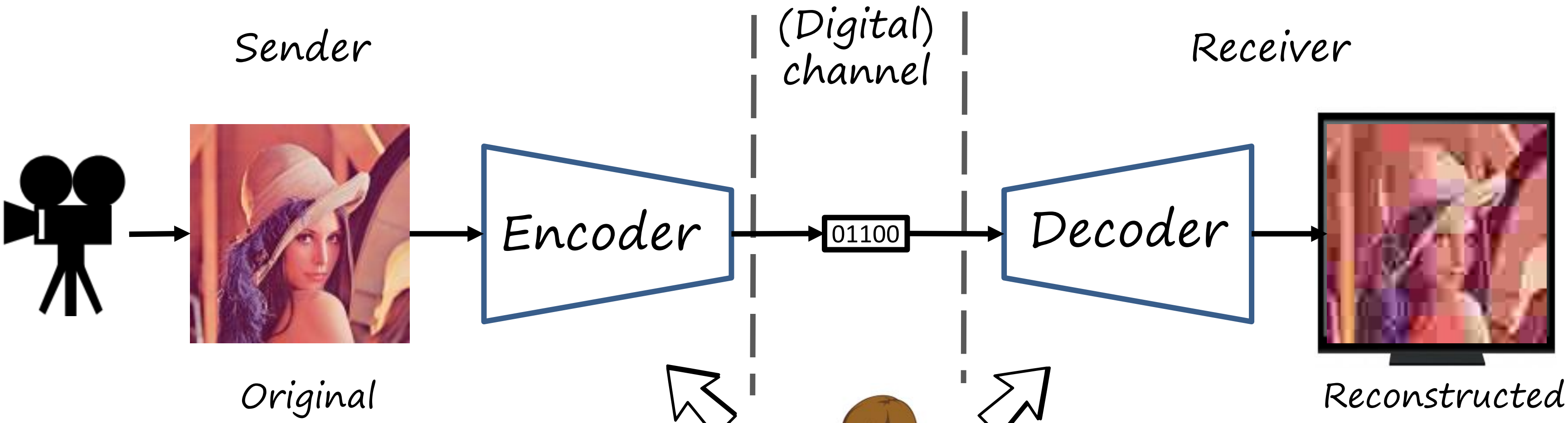
DANICE: Domain adaptation without forgetting in neural image compression

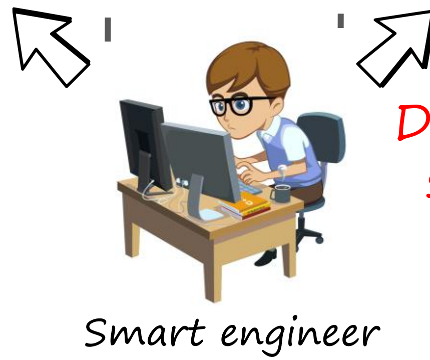
Sudeep Katakol[1,(2)], Luis Herranz[2,3], Fei Yang[2,3,4], Marta Mrak[5]

[1]University of Michigan, Ann Arbor, [2]Computer Vision Center, [3]Universitat Autònoma de Barcelona, [4]Northwestern Politechnical University, [5]BBC R&D

CLIC 2021 (@CVPR 2021)

Developing traditional image/video codecs





Design (handcrafted) syntax and coding tools

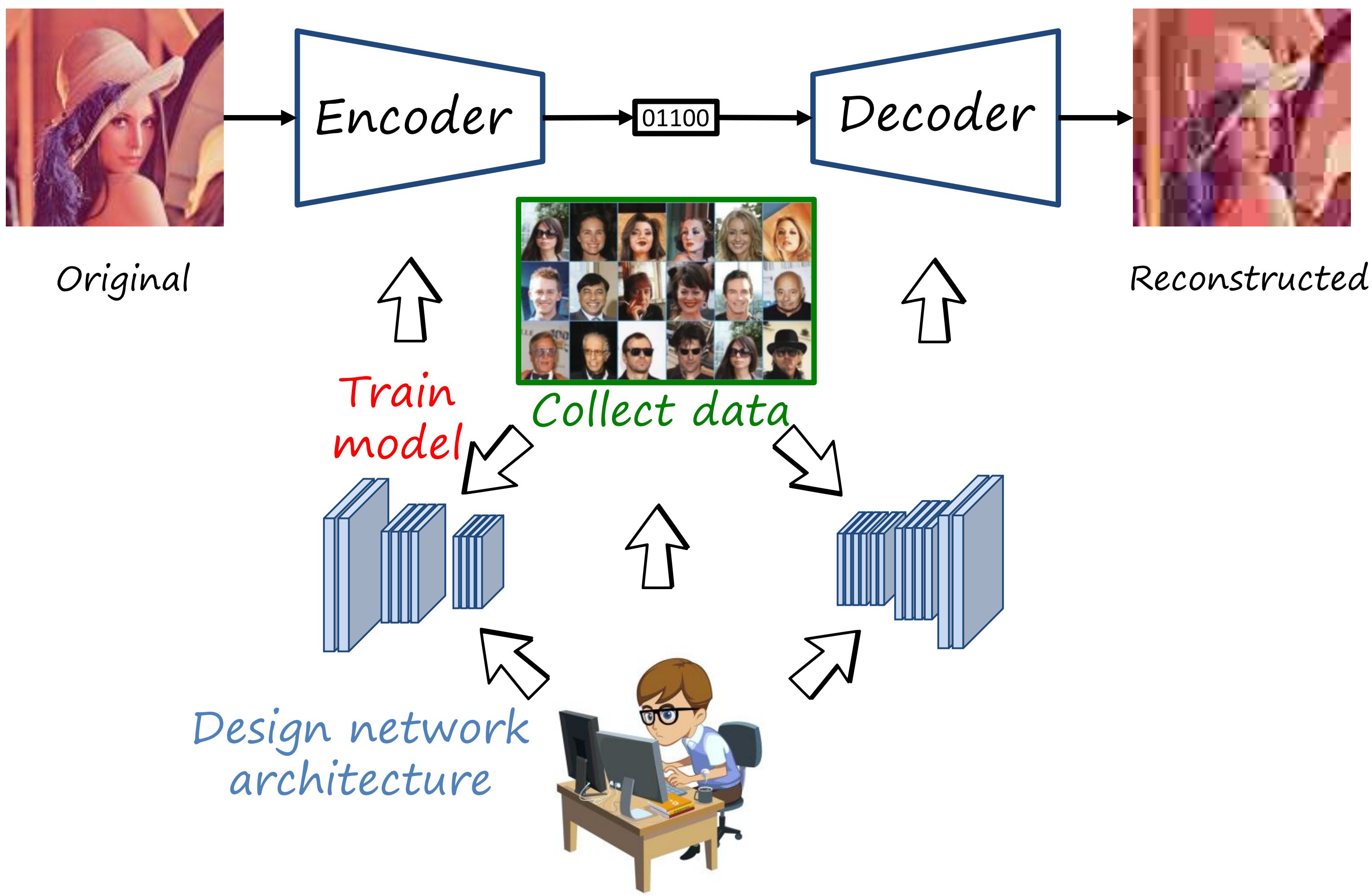
Main objectives
- Minimize rate
- Minimize distortion

Practical objectives
- Variable rate
- Low memory
- Low computation
- Low latency

Other practical considerations
- Domain-specific codecs (e.g. videoconf., screencast)
- Backward/forward compatibility with legacy formats and encoders/decoders

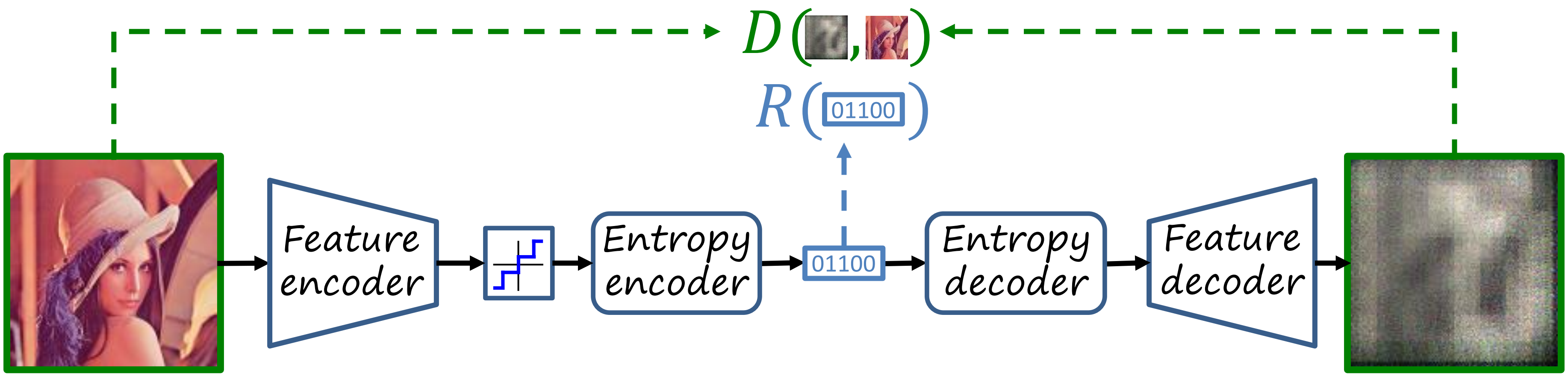
Neural image/video codecs

- Coding tools and syntax are **parametric** and **learned**
- Encoders/decoders are *deep neural networks*



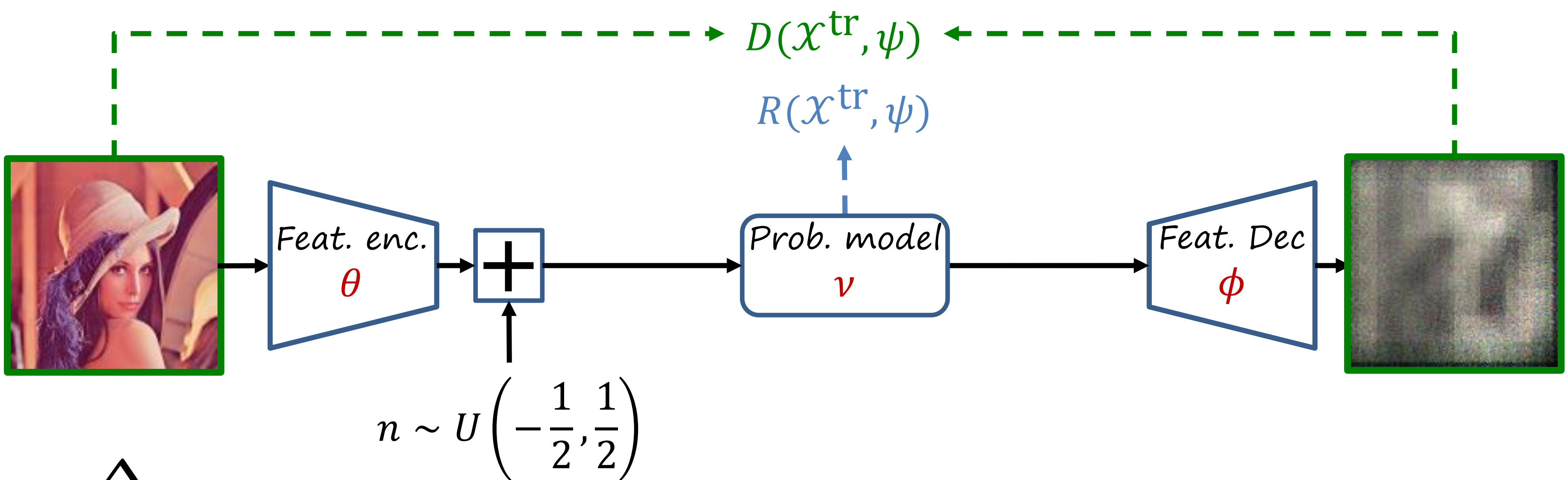
Neural image compression

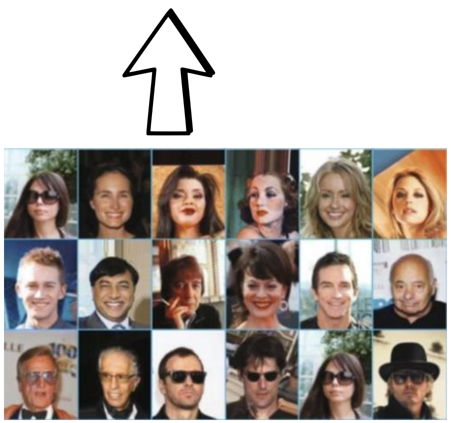
Compressive autoencoder (CAE) [Theis2017, Balle2017]
(autoencoder+quantization+entropy coding)



Neural image compression

Training (using differentiable proxies)





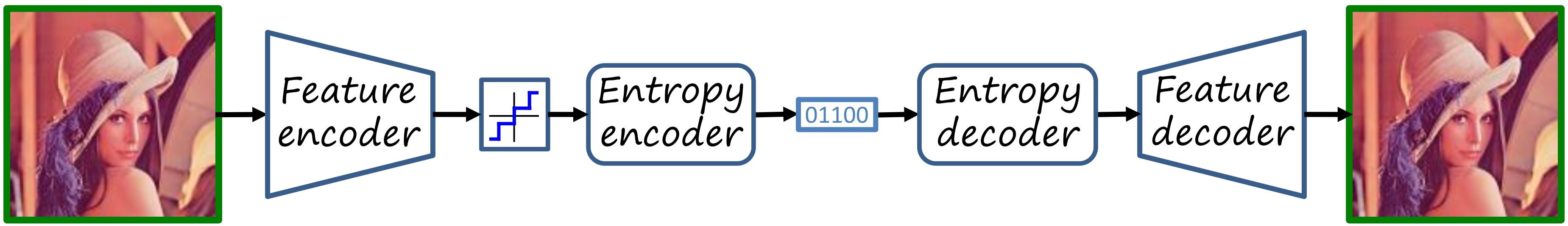
Training data $\mathcal{X}^{\text{tr}}$

Model parameters $\psi = (\theta, \phi, \nu)$

Loss $J(\mathcal{X}^{\text{tr}}, \psi; \lambda) = R(\mathcal{X}^{\text{tr}}, \psi) + \lambda D(\mathcal{X}^{\text{tr}}, \psi)$

Optimization problem $\psi^* = \min_{\psi} J(\mathcal{X}^{\text{tr}}, \psi; \lambda)$

Practical considerations in neural image compression



Main objectives
- Minimize rate
- Minimize distortion

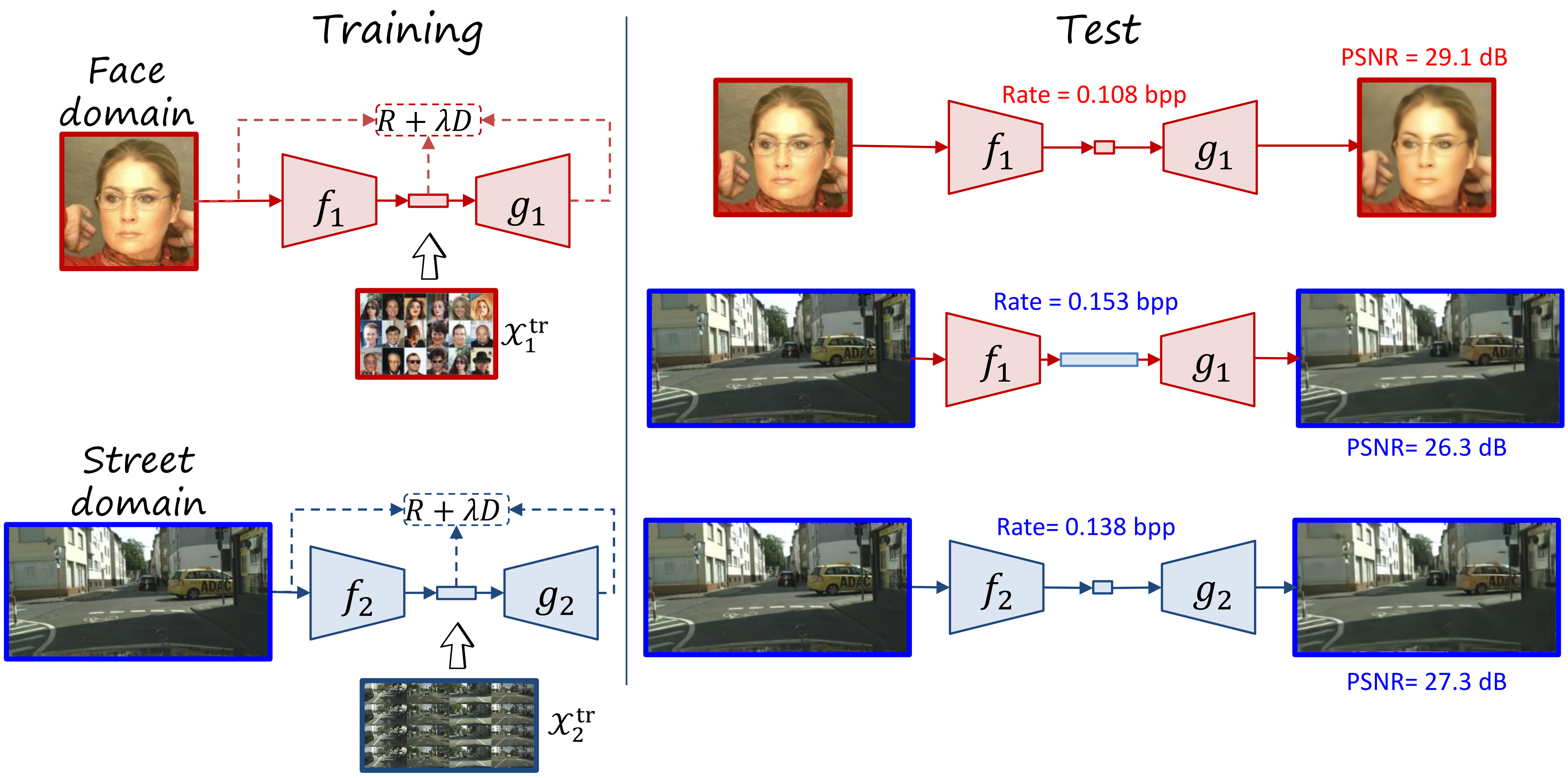
Practical objectives
- Variable rate
- Low memory
- Low computation
- Low latency

Check our paper
SlimCAE [CVPR2021]

Other practical considerations (**this work**)
- **Domain-specific codecs** (e.g. videoconference, screencast)
- **Backward/forward compatibility** (with legacy formats and encoders/decoders)

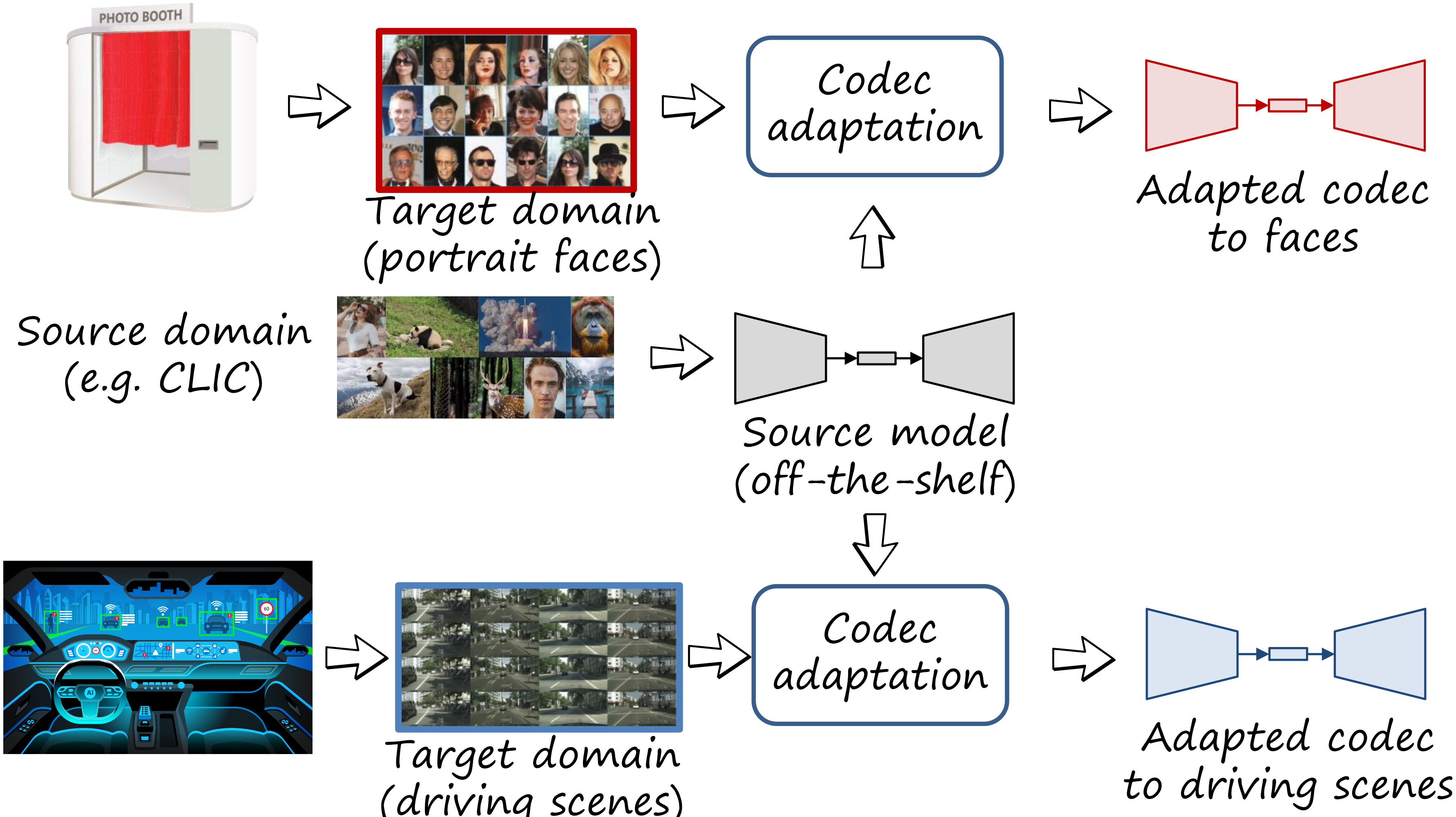
Rate-distortion optimality of learned codecs

Learned codecs are only optimal in the domain of the training data

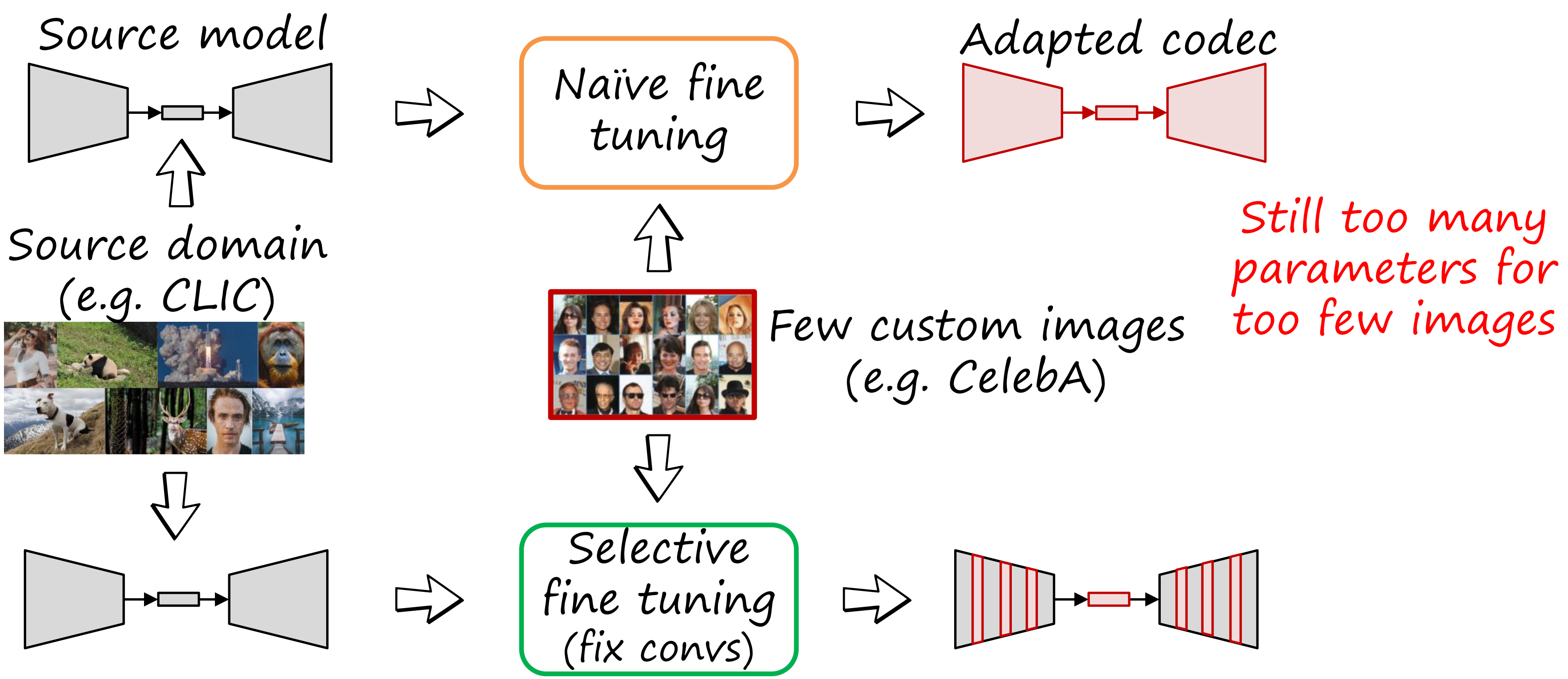


Domain Adaptation in Neural ImageomprEssion (DANICE)

Learned codecs can be customized with user content to specific domains
Problem: usually we don't have enough custom data; training is expensive
Solution: transfer pre-trained codecs



Domain adaptation via fine tuning

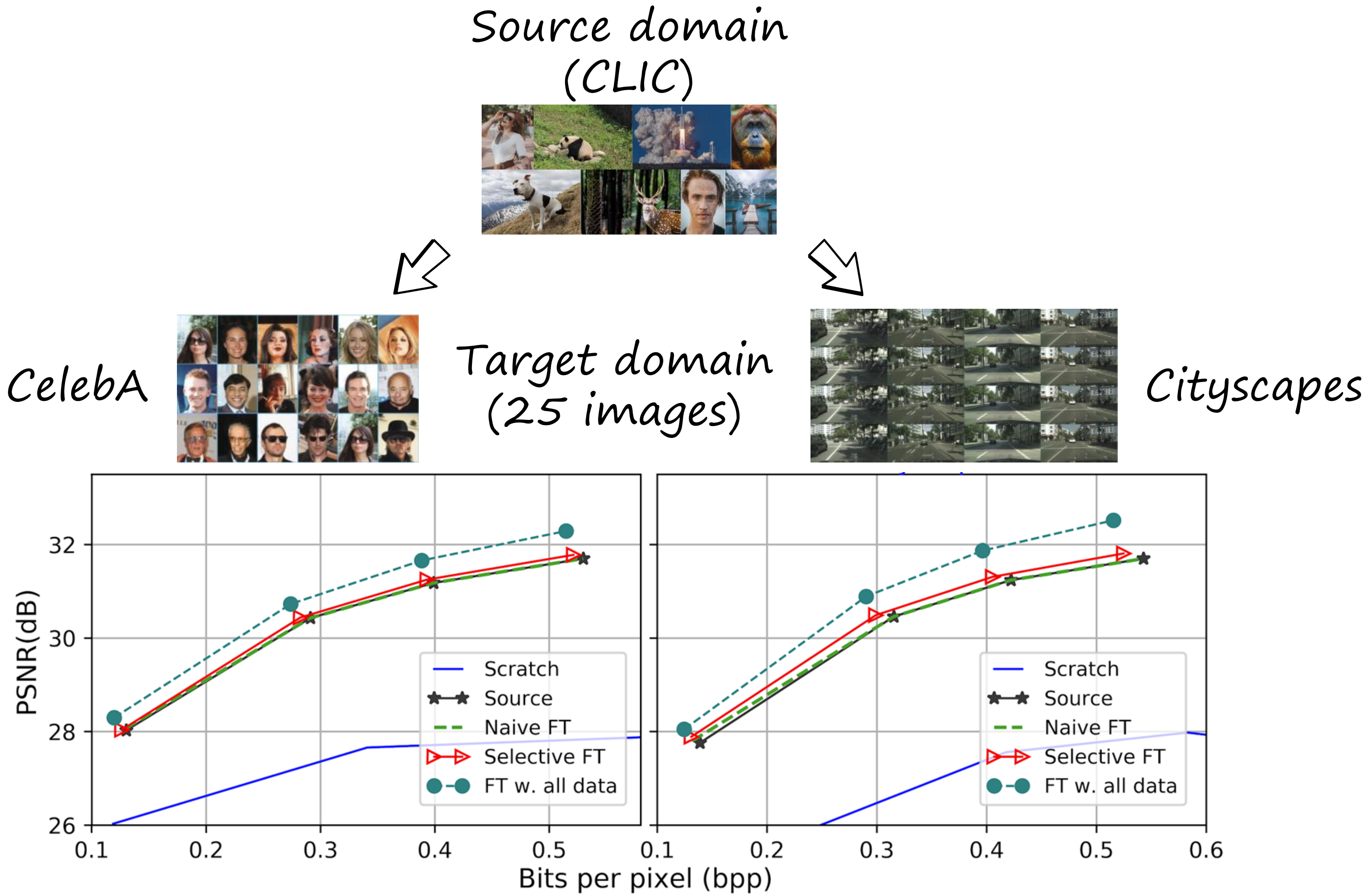


Experiments

	CLIC → CelebA		CLIC → Cityscapes	
Source model	19.24		23.93	
Number of target images	Naïve fine tuning	Selective fine tuning	Naïve fine tuning	Selective fine tuning
10	19.24	**16.46**	22.96	**17.54**
25	18.76	**14.93**	18.44	**15.79**
50	15.59	**13.73**	16.29	**15.33**

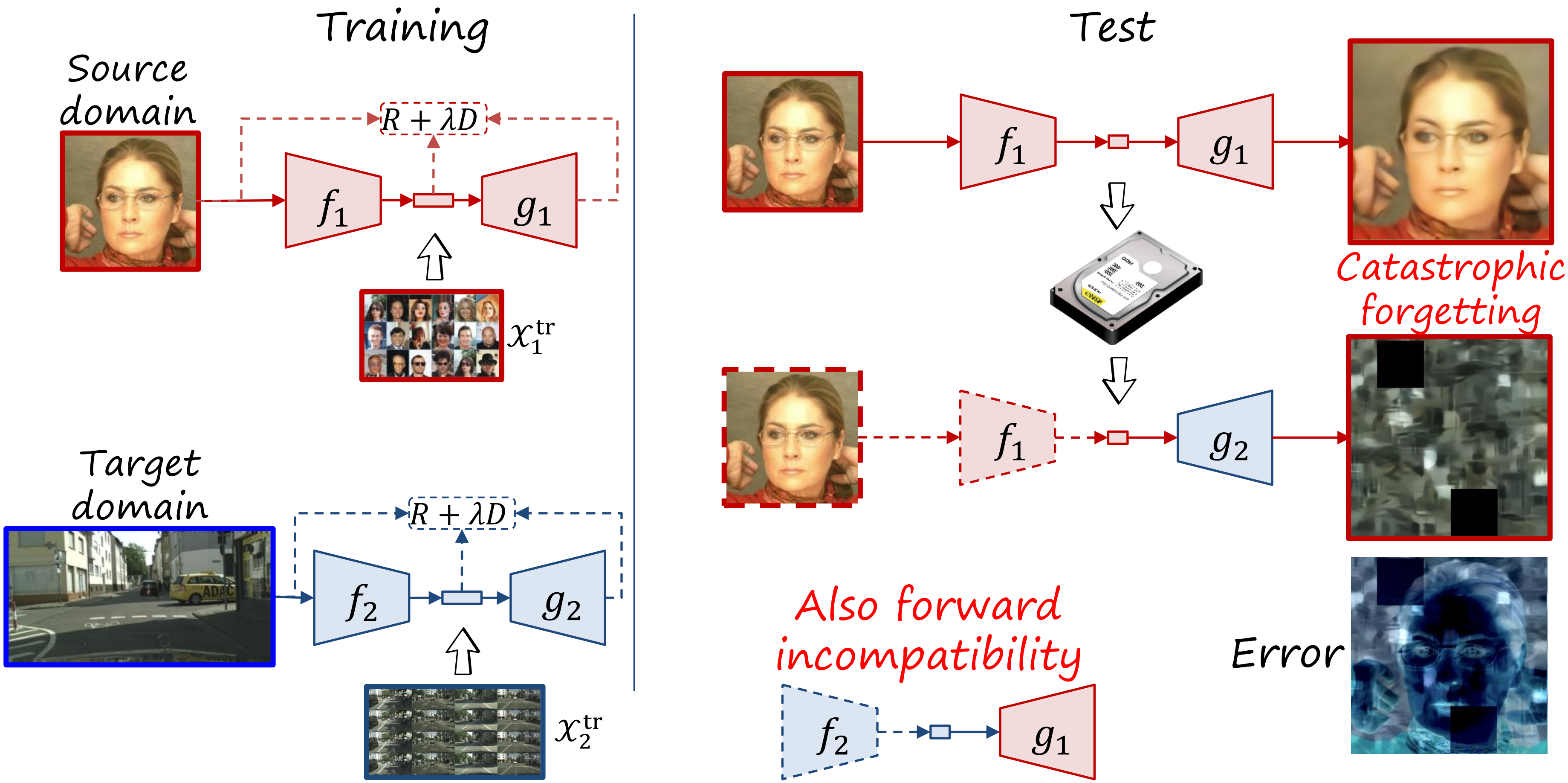
BD-rate
(reference: training with all target data)

Domain adaptation via fine tuning



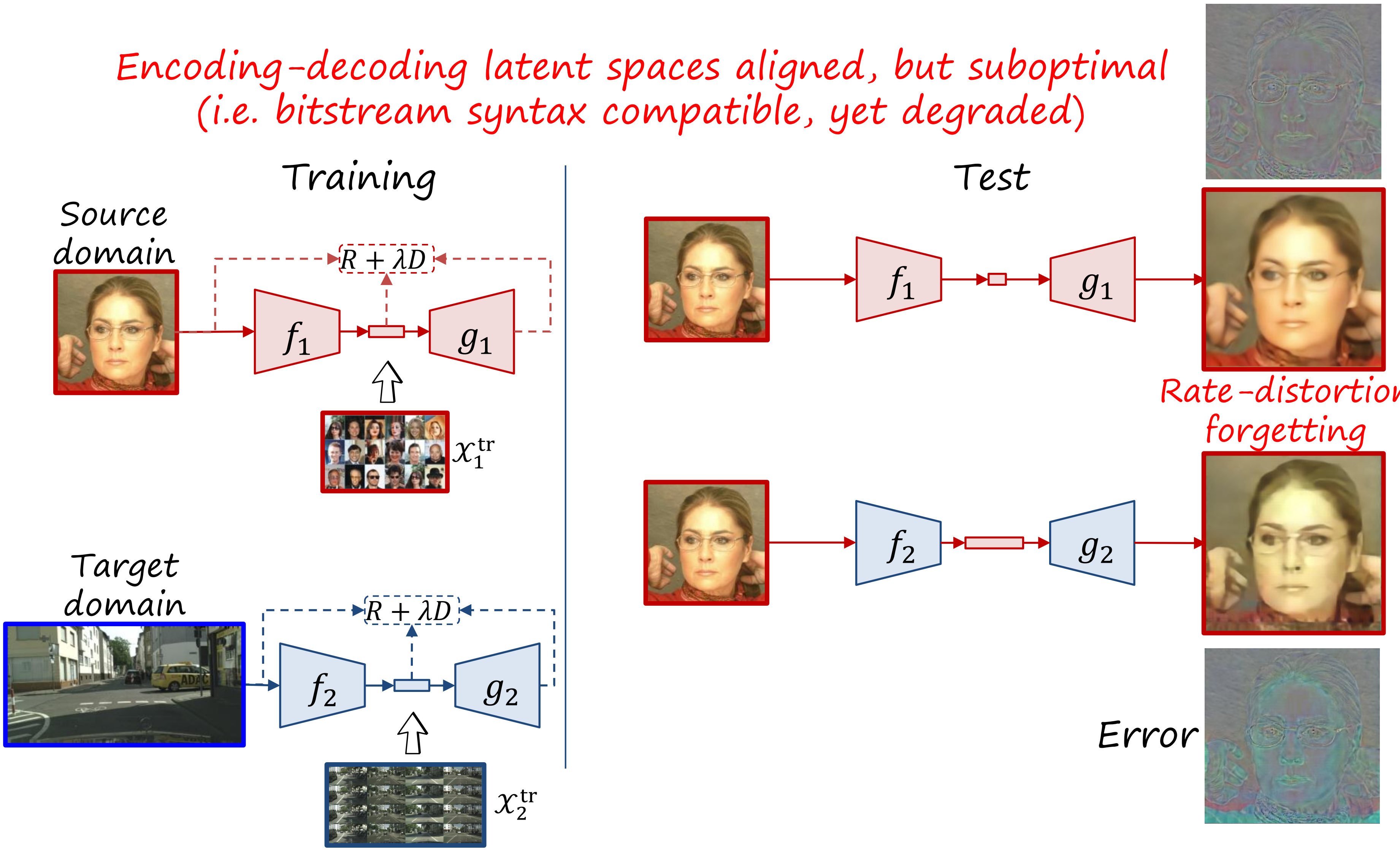
Backward incompatibility with legacy bitstreams: catastrophic forgetting

Misalignment between encoding-decoding latent spaces
(i.e. bitstream syntax incompatible)



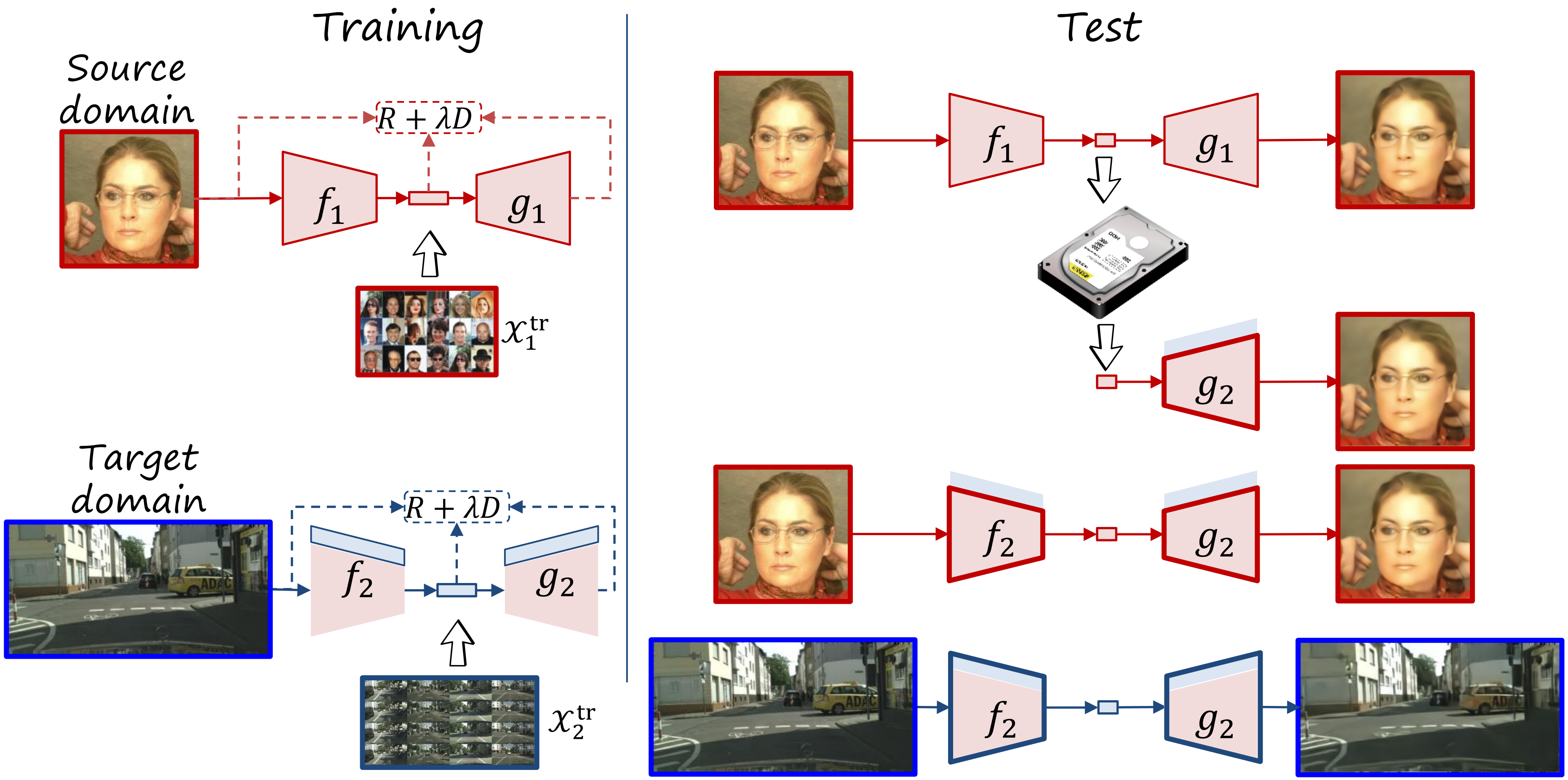
Rate-distortion forgetting

Encoding-decoding latent spaces aligned, but suboptimal
(i.e. bitstream syntax compatible, yet degraded)



Codec adaptation without forgetting (CAwF)

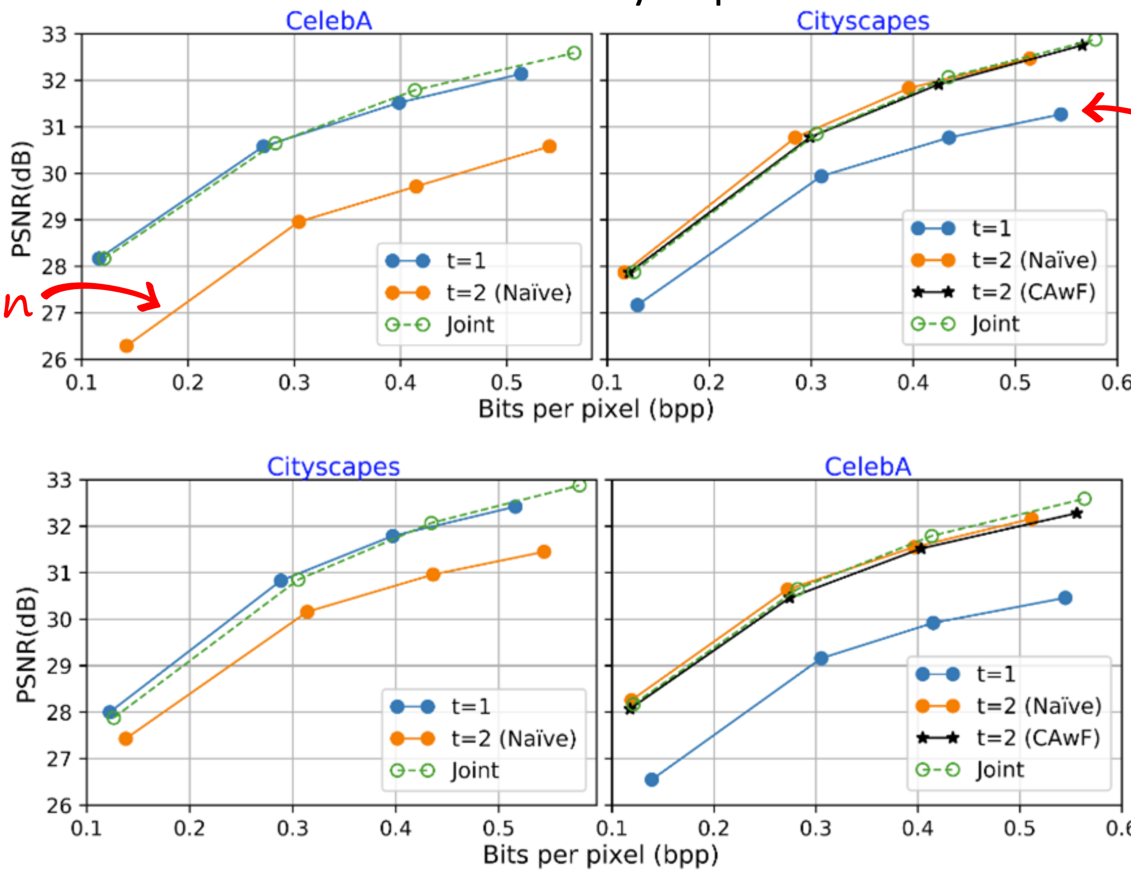
Freeze source codec, and learn target codec as an enhancement layer
Drawback: adds additional parameters



Codec adaptation without forgetting (CAwF)

CelebA→Cityscapes

Naïve adaptation forgets at t=2

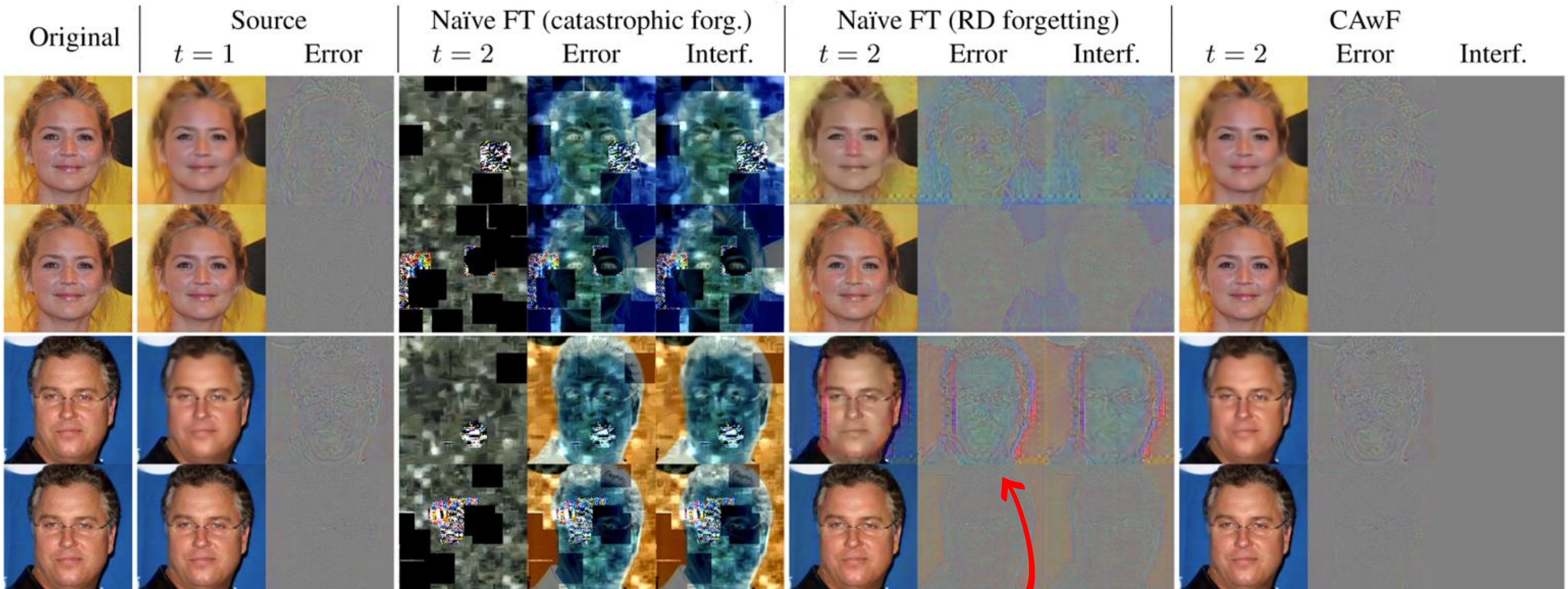


Source codec is suboptimal at t=2

Cityscapes→CelebA

Codec adaptation without forgetting (CAwF)

CelebA→Cityscapes
(source domain)



Adaptation
artifacts

Thanks!

https://arxiv.org/abs/2103.15726

Sudeep Katakol

Luis Herranz

Fei Yang

Marta Mrak

UAB Universitat Autònoma de Barcelona

BBC | Research & Development