Practical image and video compression with deep neural networks

Luis Herranz

Computer Vision Center
Universitat Autònoma de Barcelona

March 2022

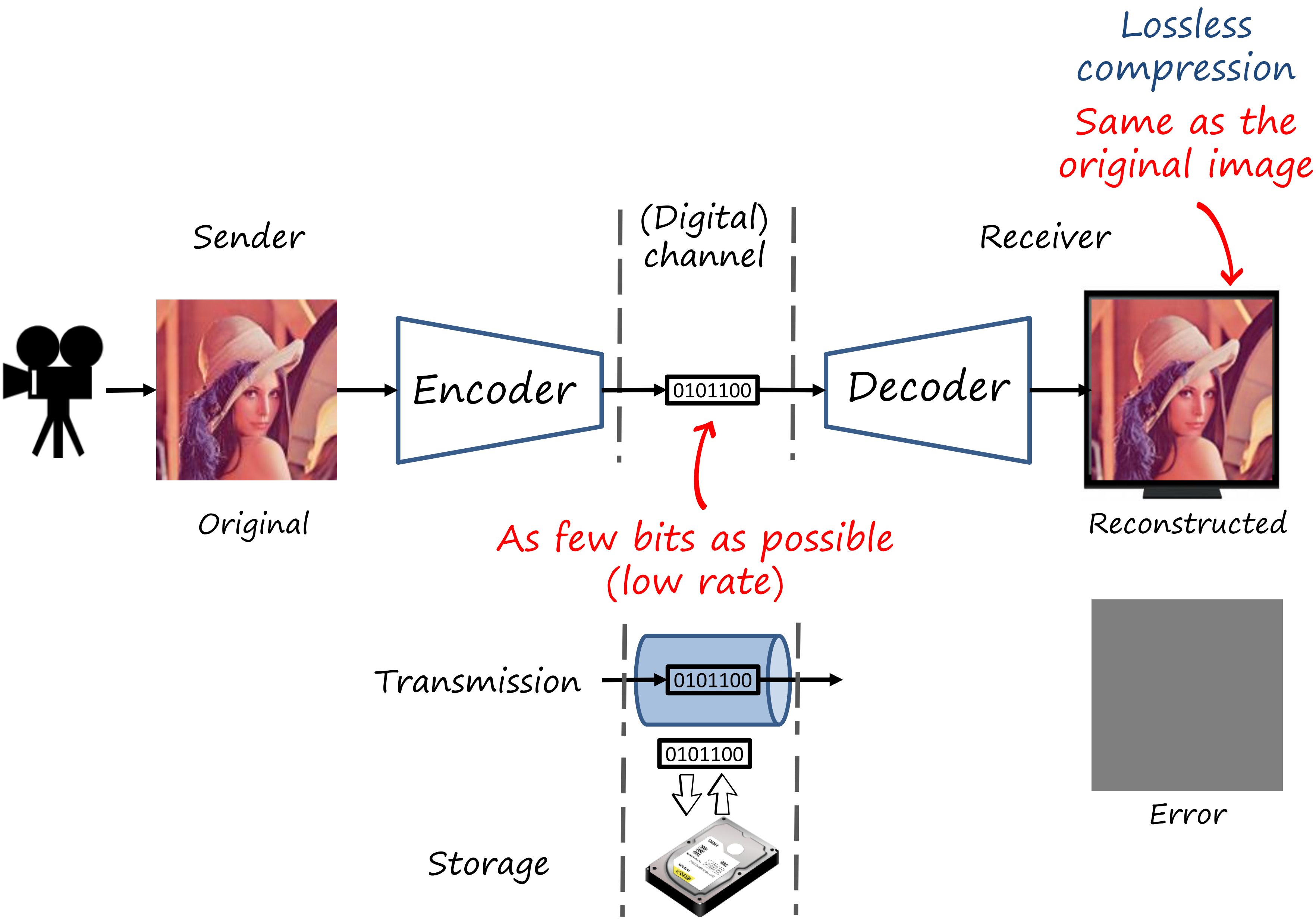
Outline

- Introduction: image/video coding
- Compression with neural networks
- Towards practical image compression
- Visual quality: perception vs distortion
- Video restoration and applications

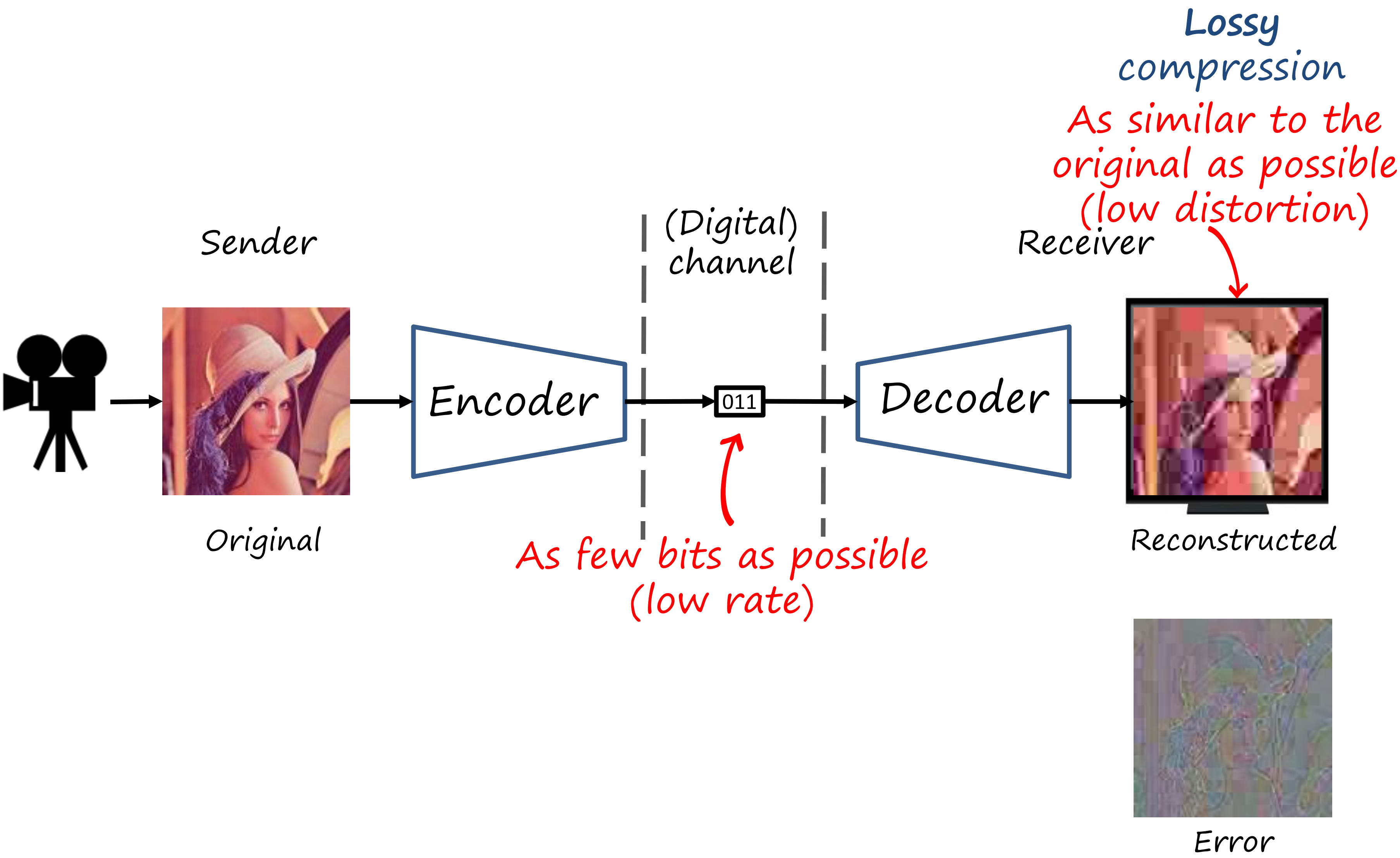
Outline

- Introduction: image/video coding
- Compression with neural networks
- Towards practical image compression
- Visual quality: perception vs distortion
- Video restoration and applications

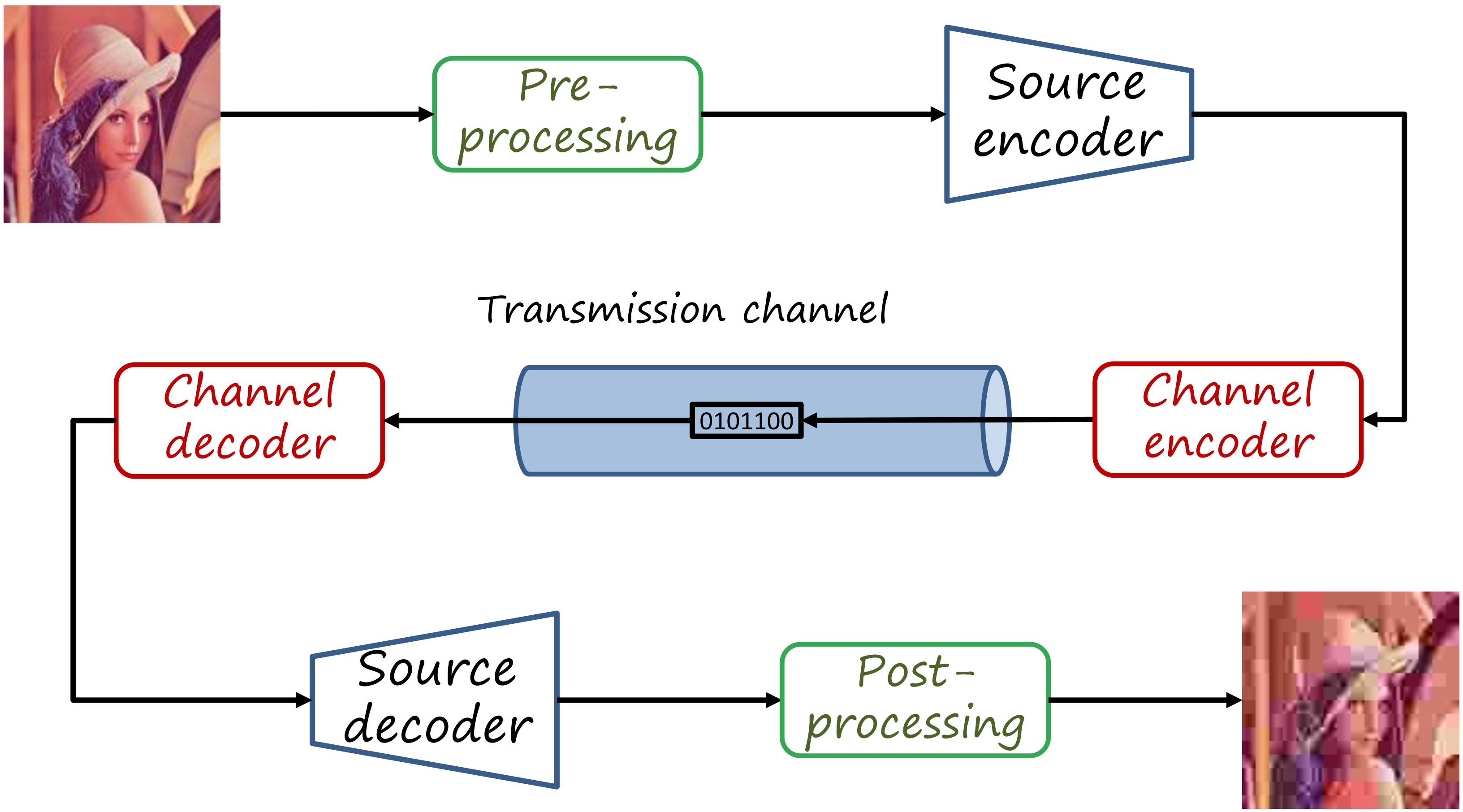
The visual communication problem



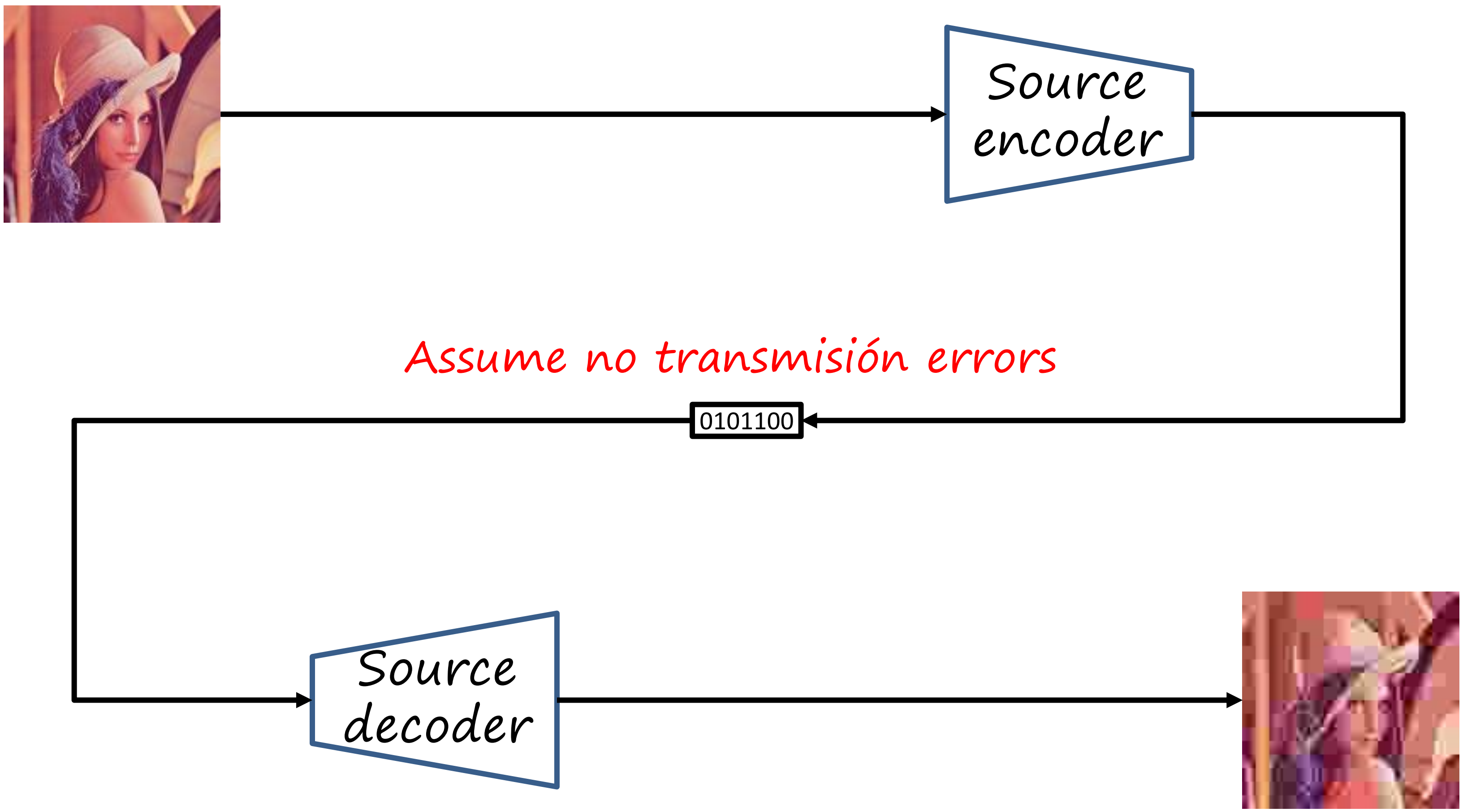
The visual communication problem



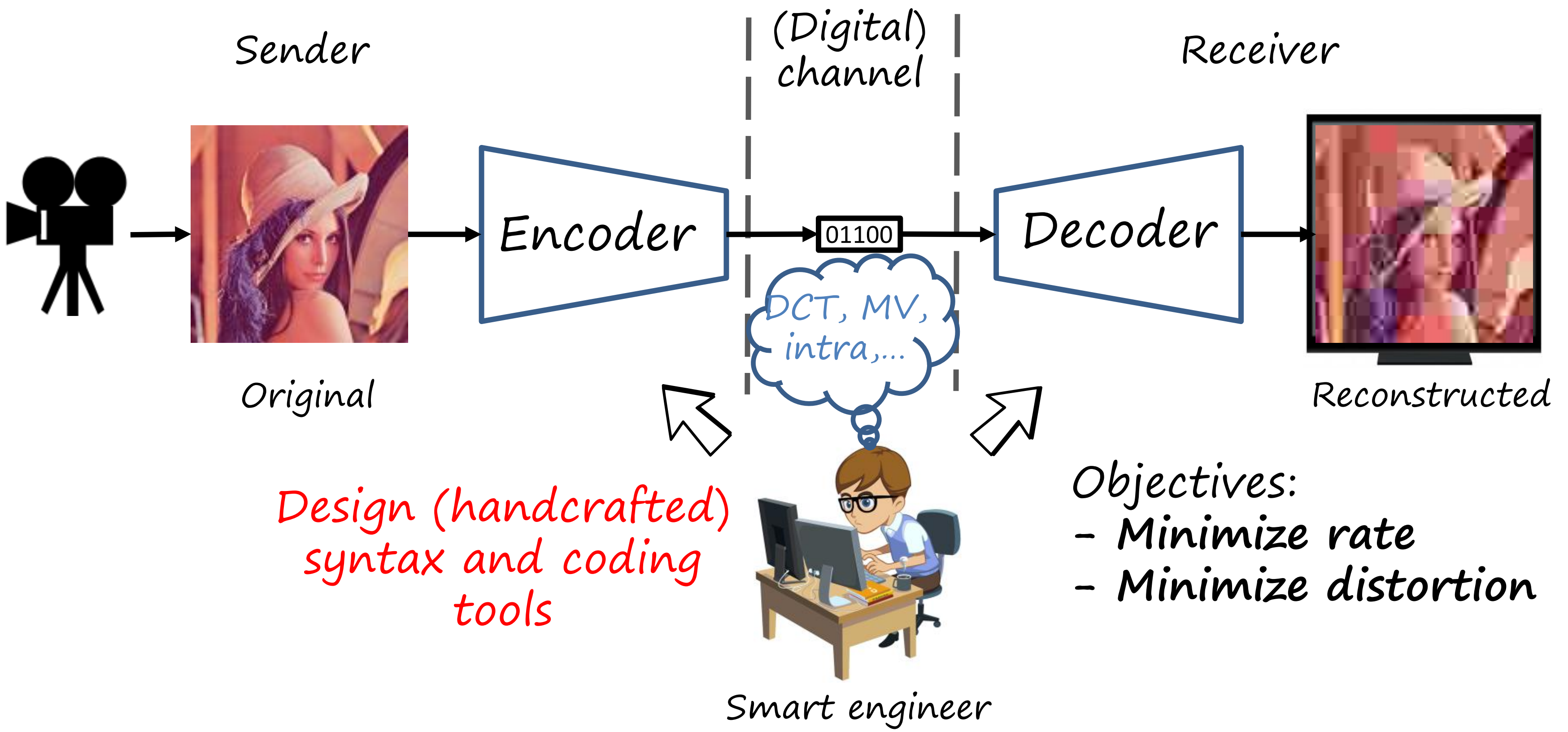
Pre/post-processing, source coding and channel coding



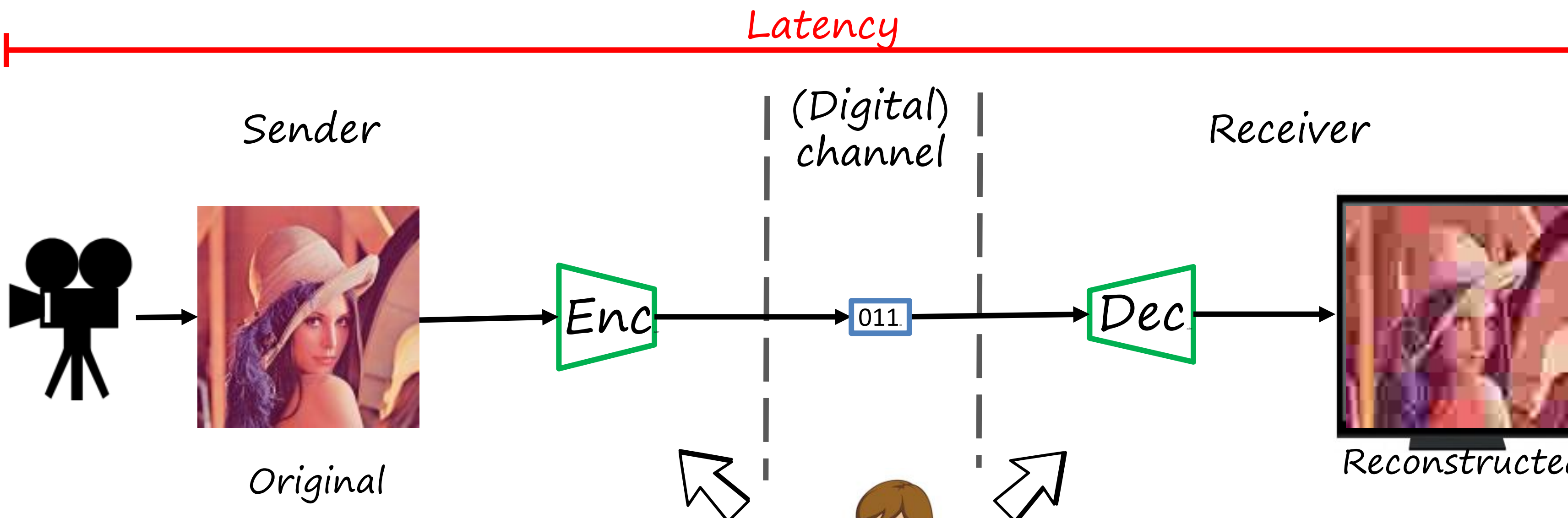
Source coding only



Developing traditional image/video codecs



… for practical applications

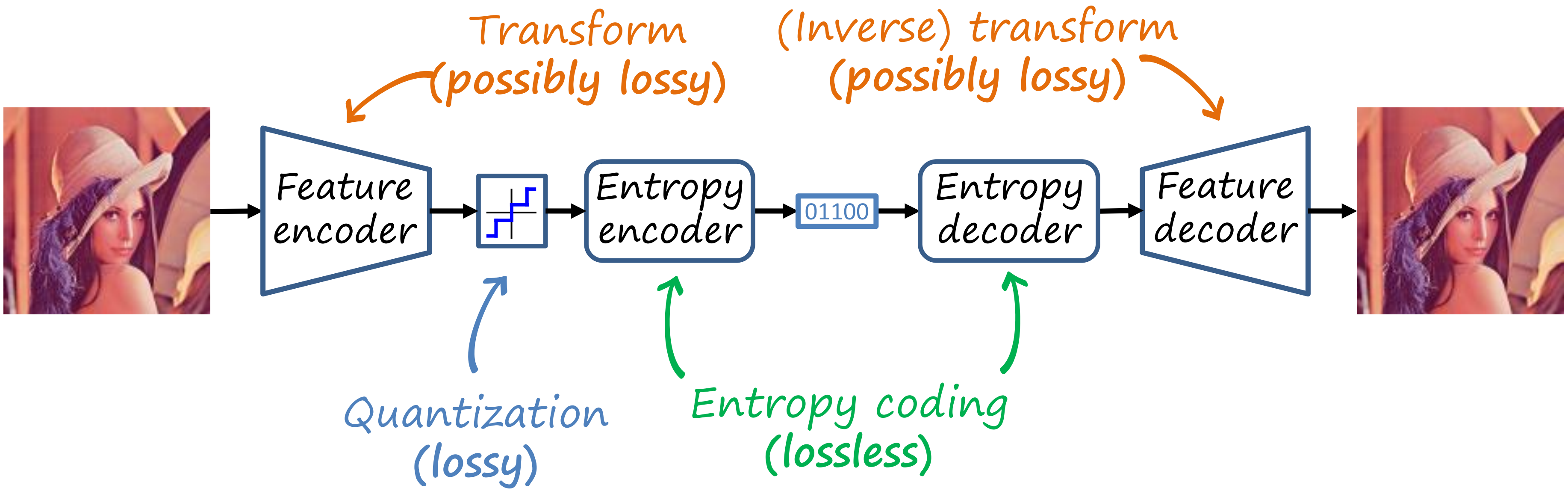


Design (handcrafted) syntax and coding tools

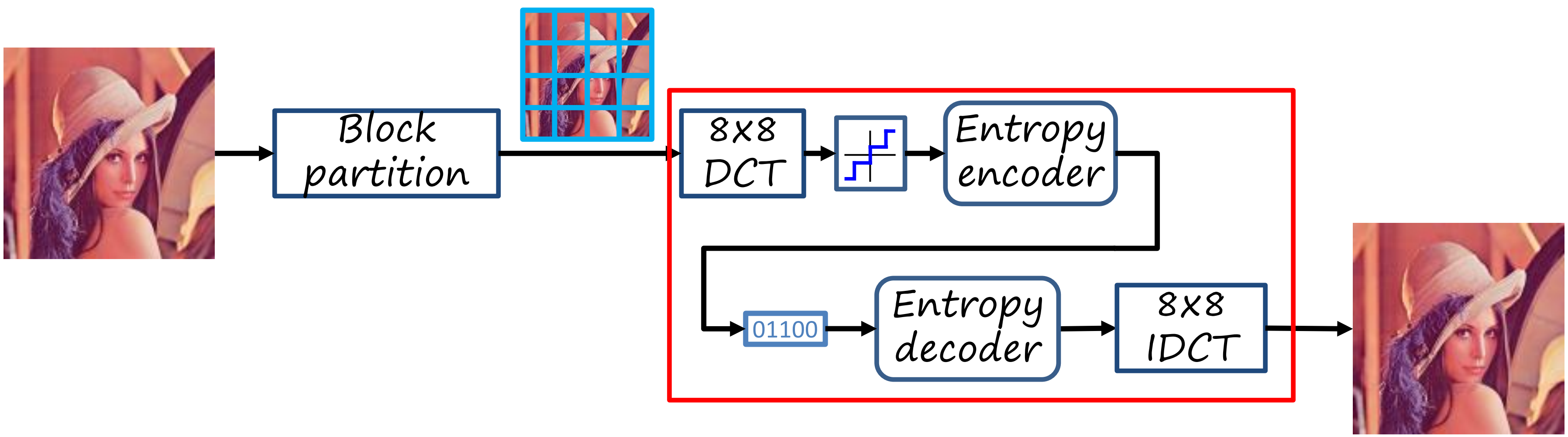
Smart engineer

Objectives:
- Minimize rate
- Minimize distortion
- Variable rate
- Low memory
- Low computation
- Low latency
- Compatibility
- Domain-specific

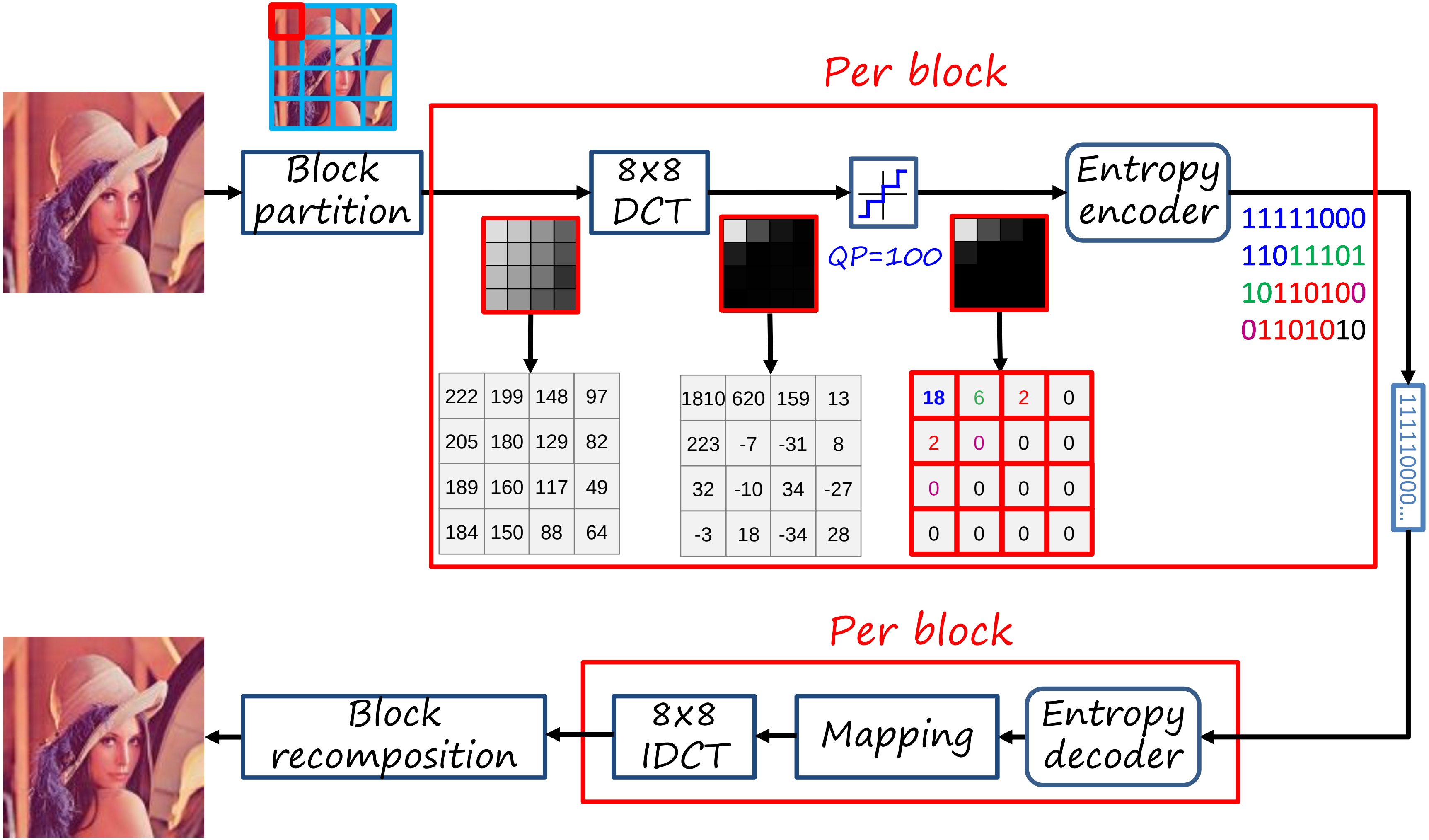
Transform coding pipeline



Example: block-based transform coding (e.g. JPEG, MPEG-2, H.264)



Transform coding pipeline: JPEG



222	199	148	97
205	180	129	82
189	160	117	49
184	150	88	64

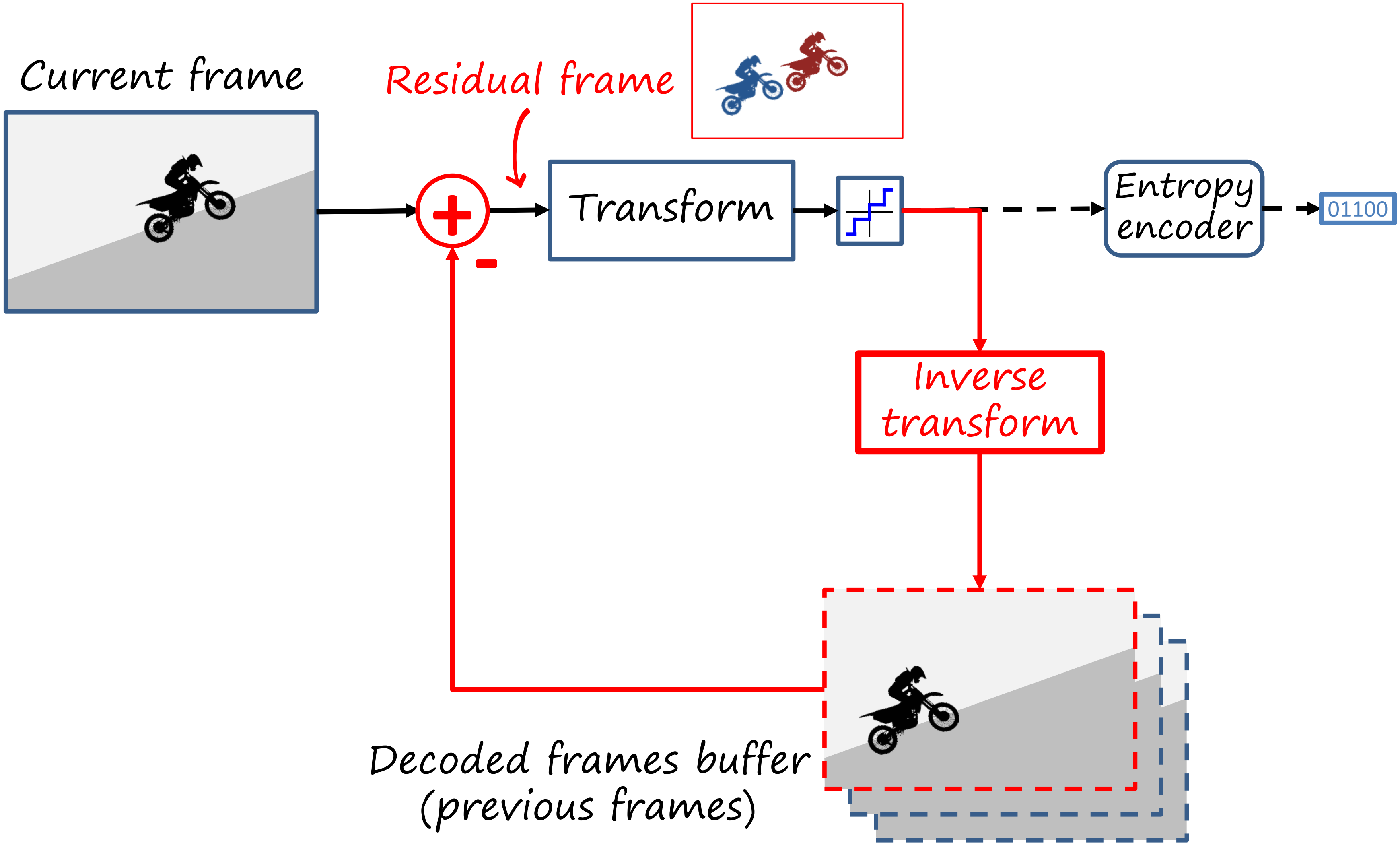
1810	620	159	13
223	-7	-31	8
32	-10	34	-27
-3	18	-34	28

18	6	2	0
2	0	0	0
0	0	0	0
0	0	0	0

Slide partly adapted from T. Wiegand

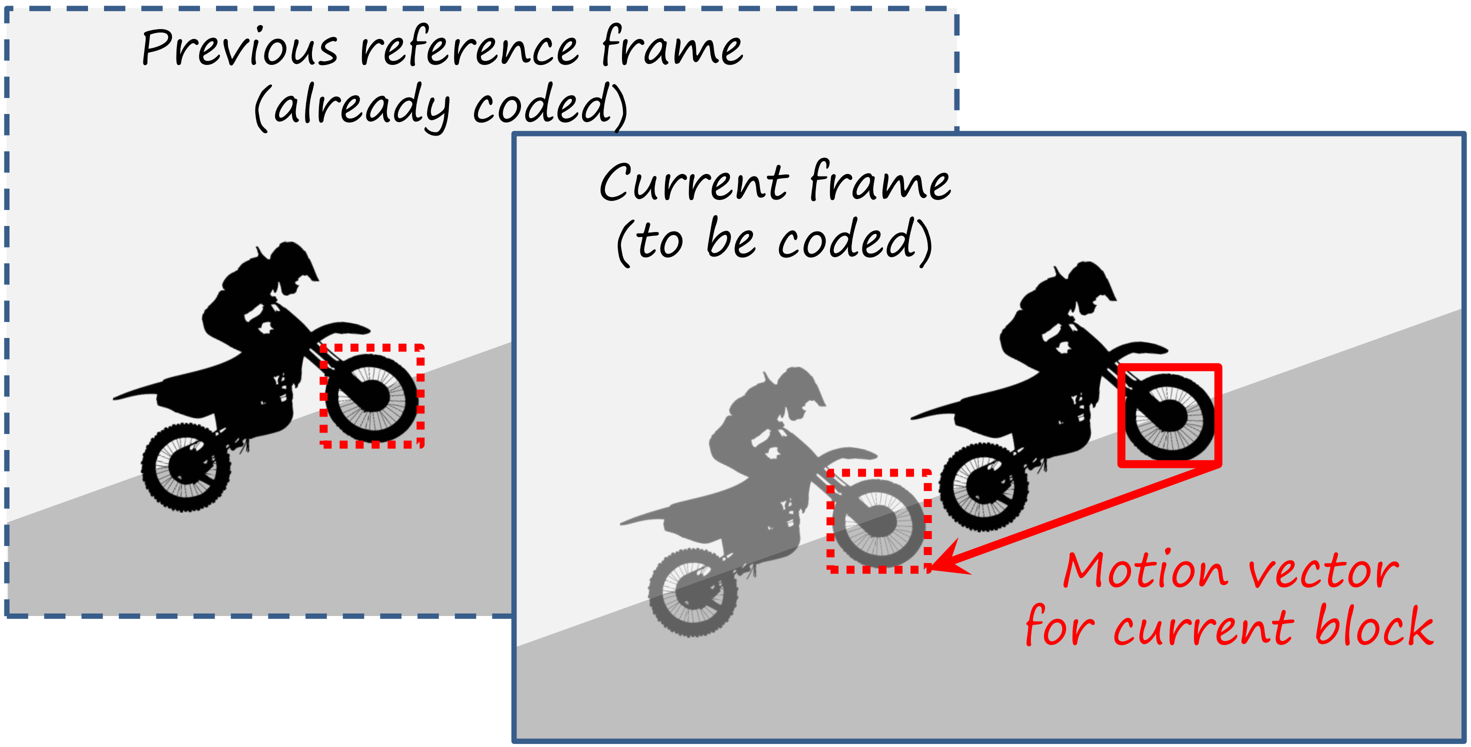
Coding video: temporal redundancy

Estimate current frame from previous coded ones



Motion-compensated prediction

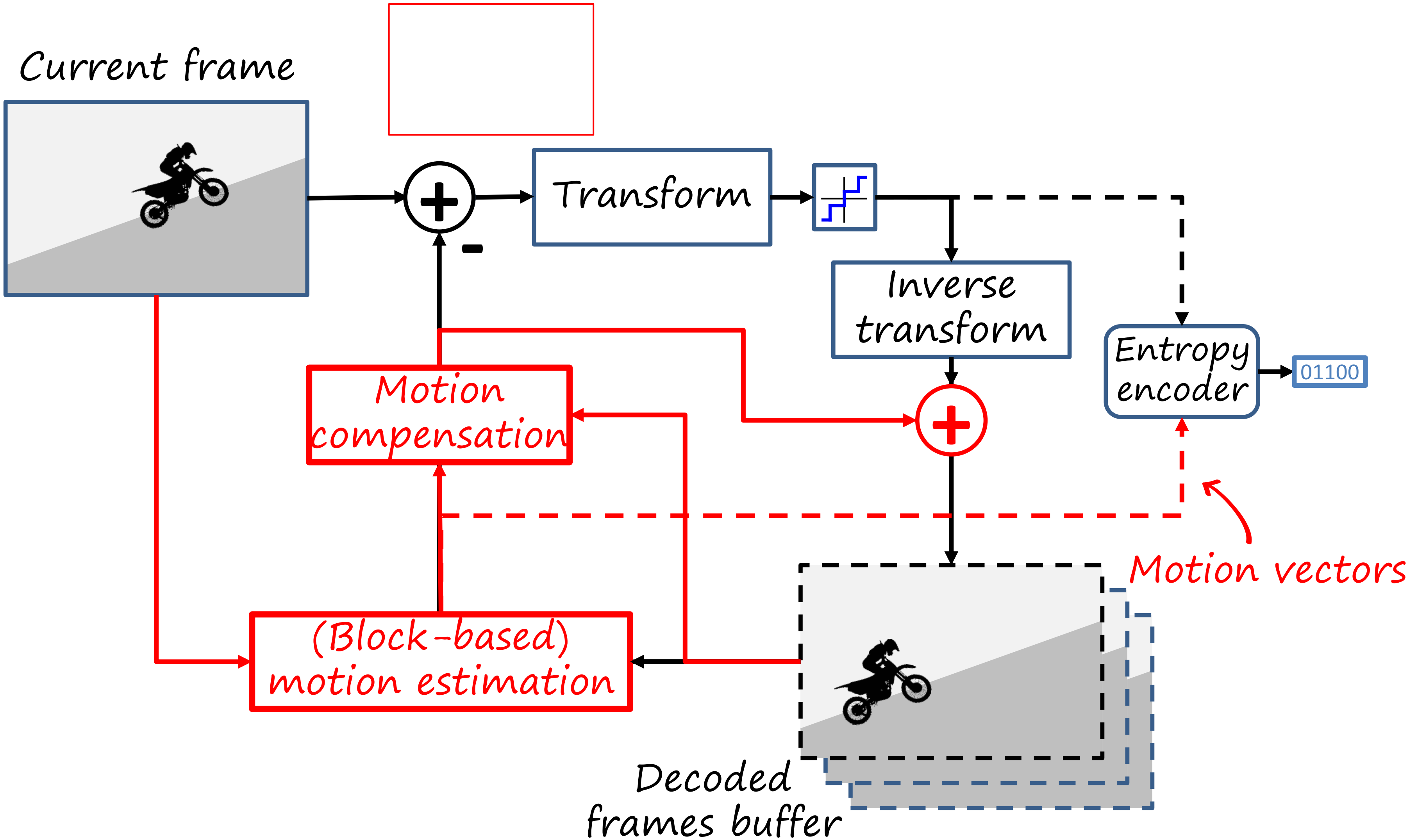
Try to align frames: find most similar blocks in the reference frame



Inspired by T. Wiegand's slide

Motion-compensated video coding

Estimate current frame from previous coded ones
Encode the motion vectors

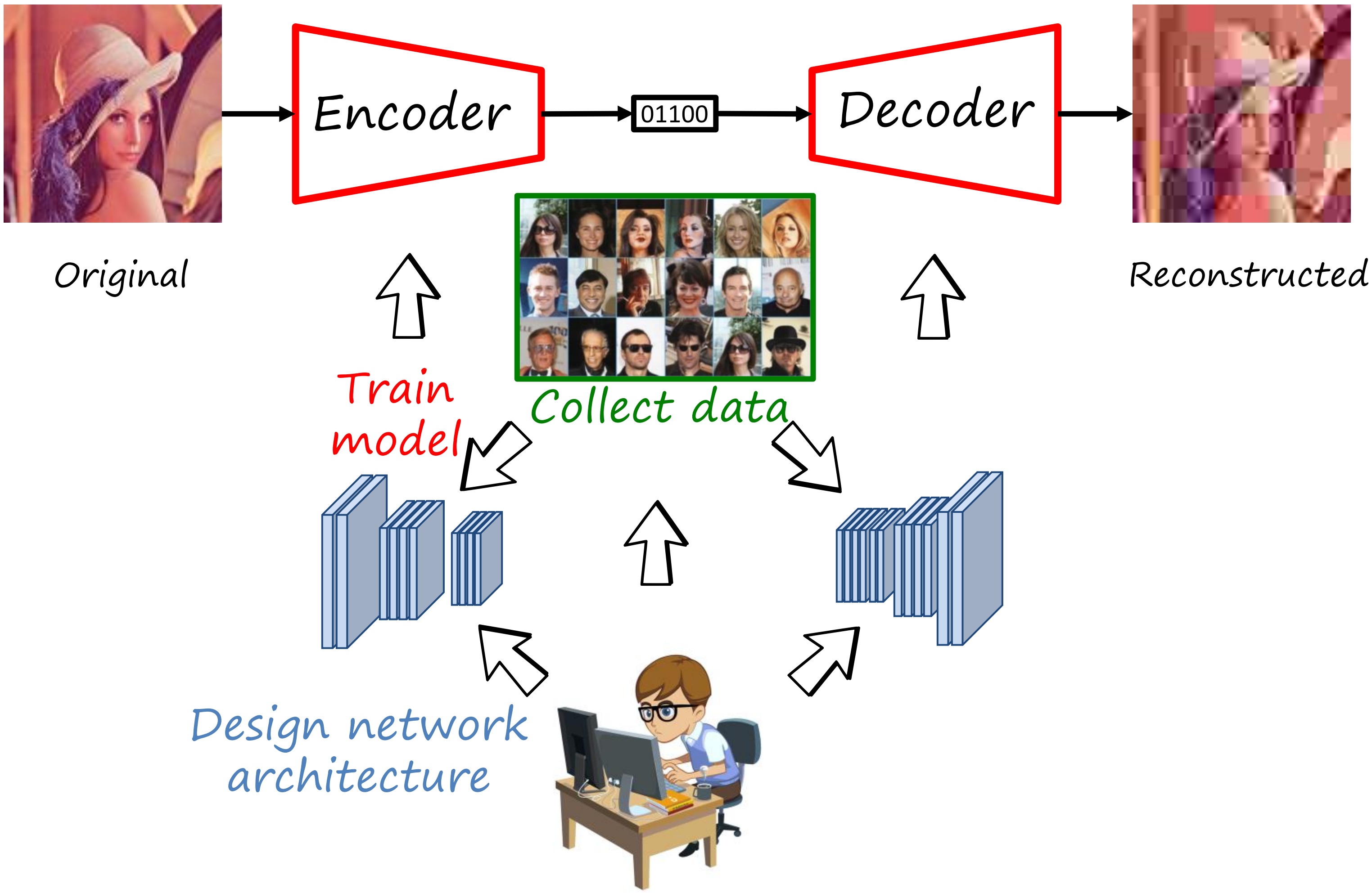


Outline

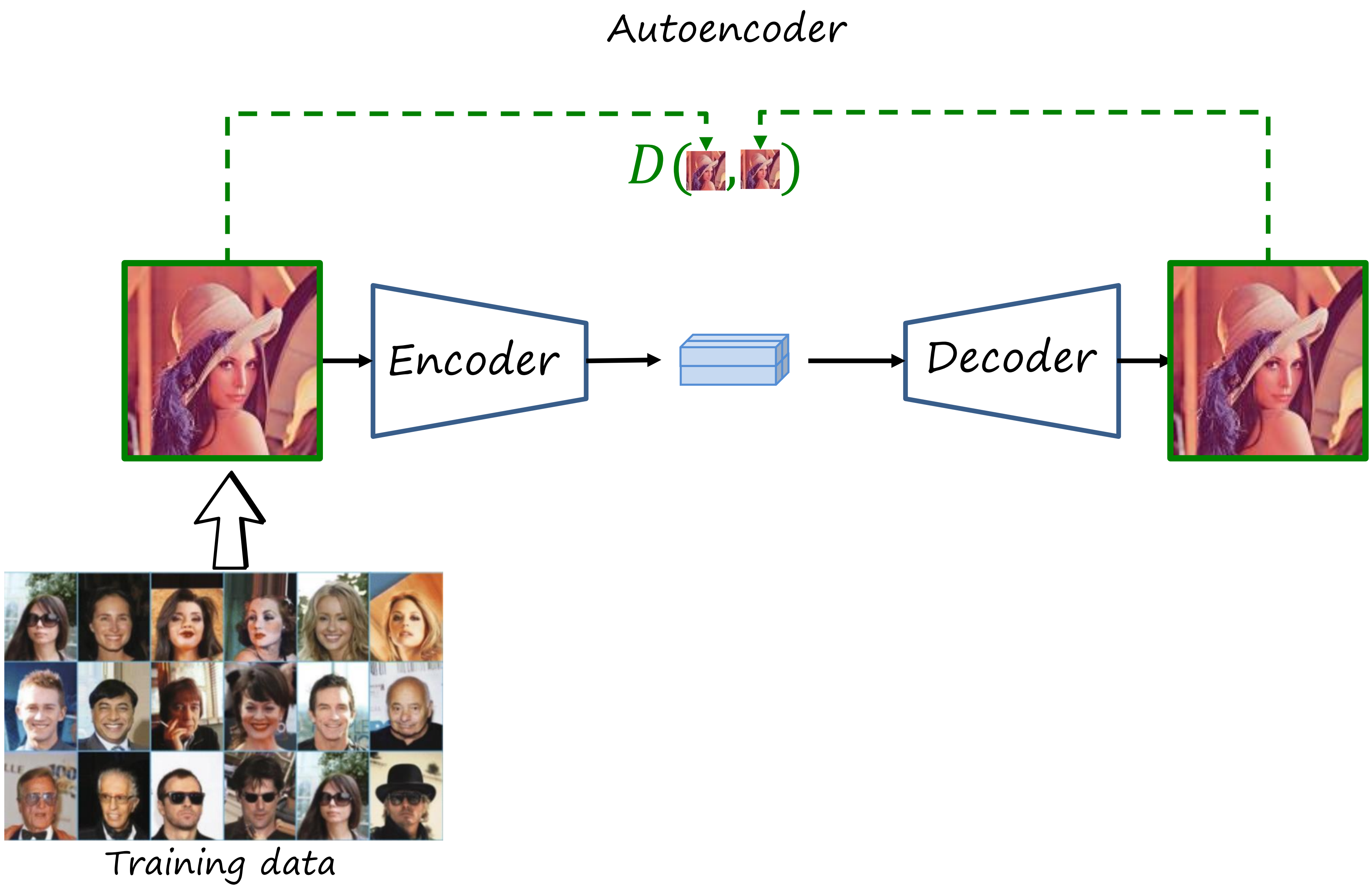
- Introduction: image/video coding
- Compression with neural networks
- Towards practical image compression
- Visual quality: perception vs distortion
- Video restoration and applications

Neural image codecs

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

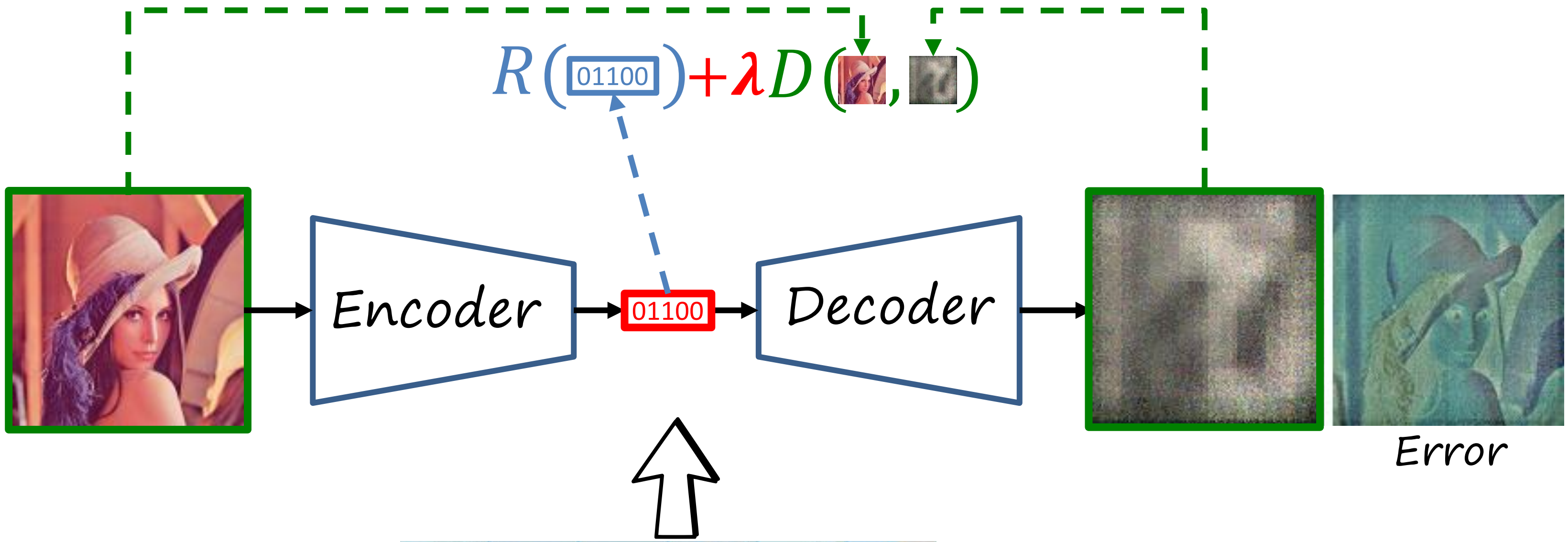


Neural image compression

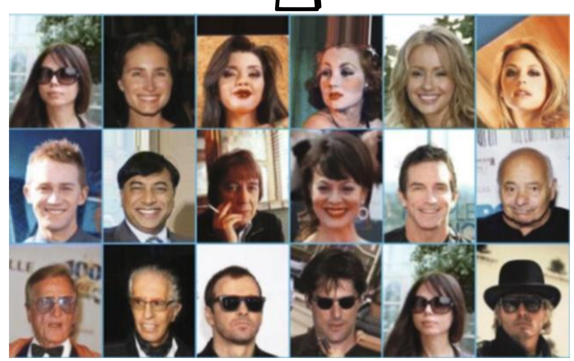


Neural image compression

Compressive autoencoder (CAE) [Theis2017, Balle2017]
(autoencoder+binary latent representation)



Optimize a weighted rate-distortion loss
(λ controls the tradeoff)



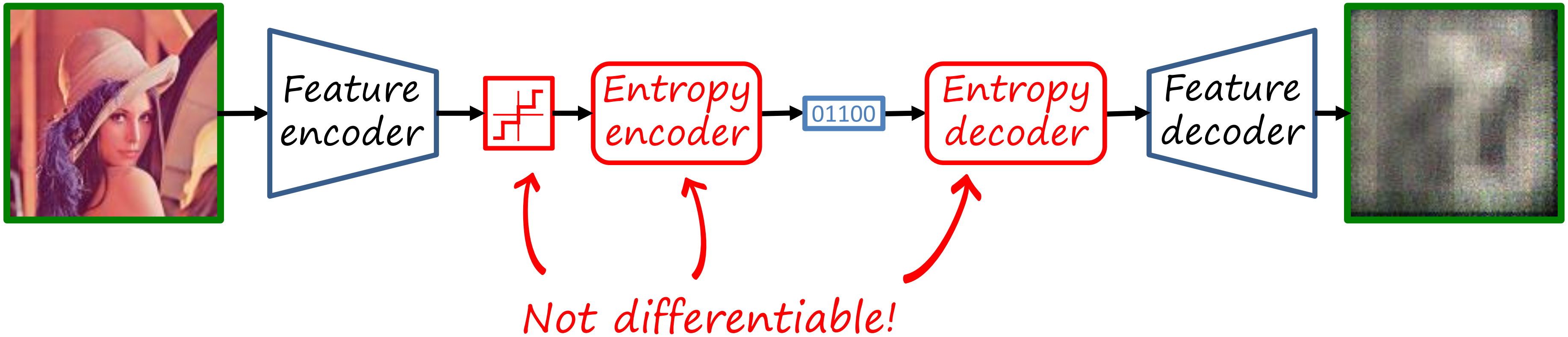
Training data

Distortion is typically mean square error (MSE)

[Balle et al. End-to-end Optimized Image Compression,](#) ICLR 2017
[Theis et al., Lossy Image Compression with Compressive Autoencoders,](#) ICLR 2017

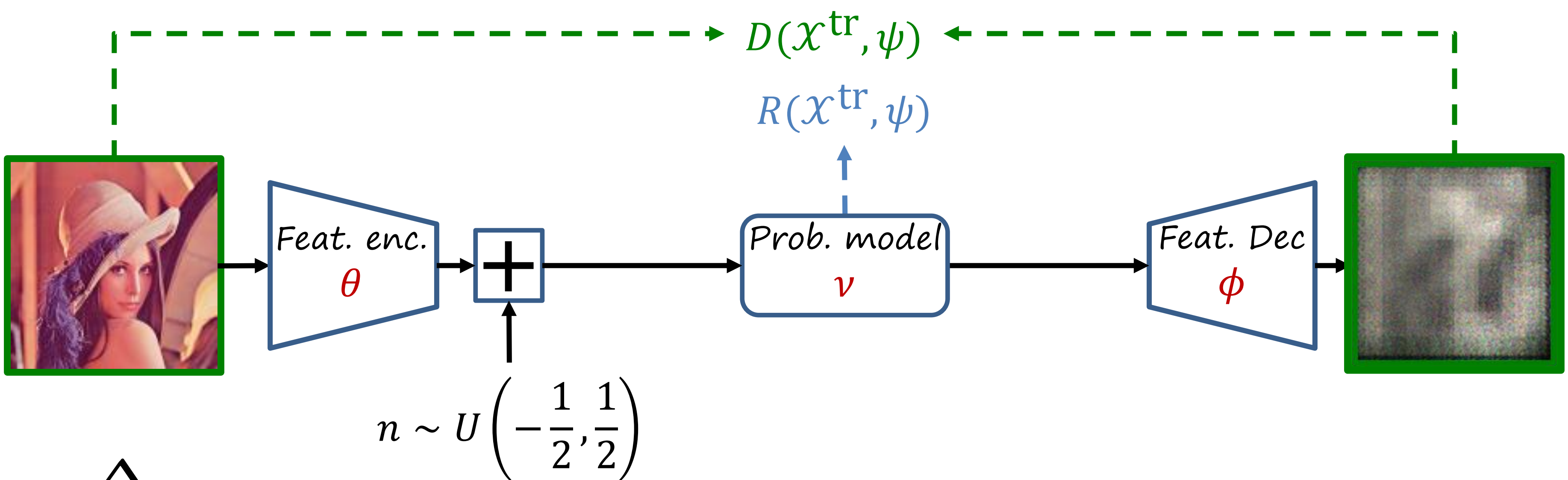
Typical pipeline

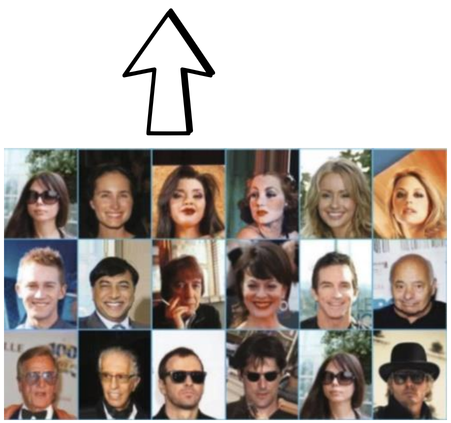
Compressive autoencoder (CAE) [Theis2017, Balle2017]
(autoencoder+quantization+entropy coding)



Architecture (training)

Use differentiable proxies for end-to-end training



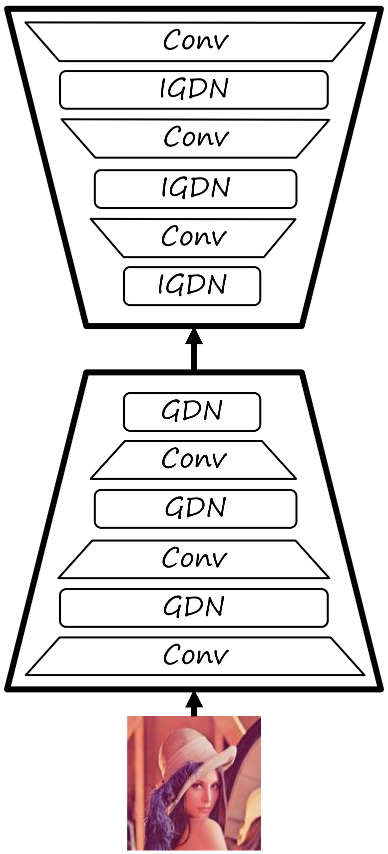


Training data $\mathcal{X}^{\mathrm{tr}}$

Model parameters	$\psi = (\theta, \phi, \nu)$
Loss	$J(\mathcal{X}^{\mathrm{tr}}, \psi; \lambda) = R(\mathcal{X}^{\mathrm{tr}}, \psi) + \lambda D(\mathcal{X}^{\mathrm{tr}}, \psi)$
Optimization problem	$\psi^* = \min_{\psi} J(\mathcal{X}^{\mathrm{tr}}, \psi; \lambda)$

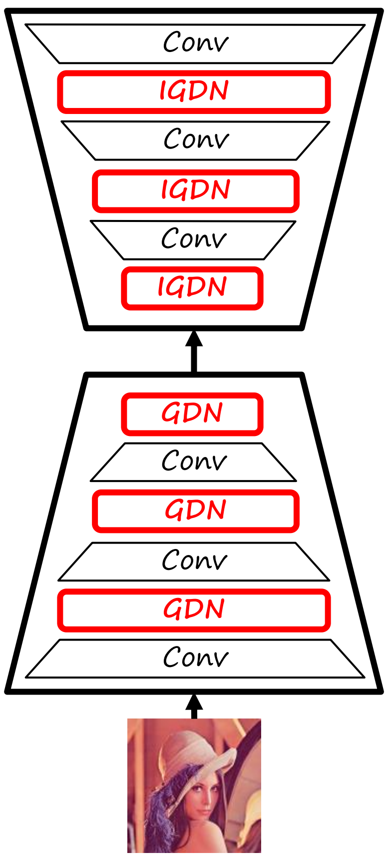
Autoencoder architecture

Balle et al.
[ICLR2017]



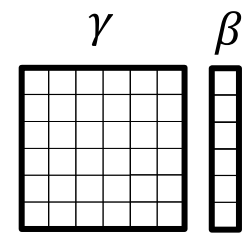
Autoencoder architecture

Balle et al.
[ICLR2017]



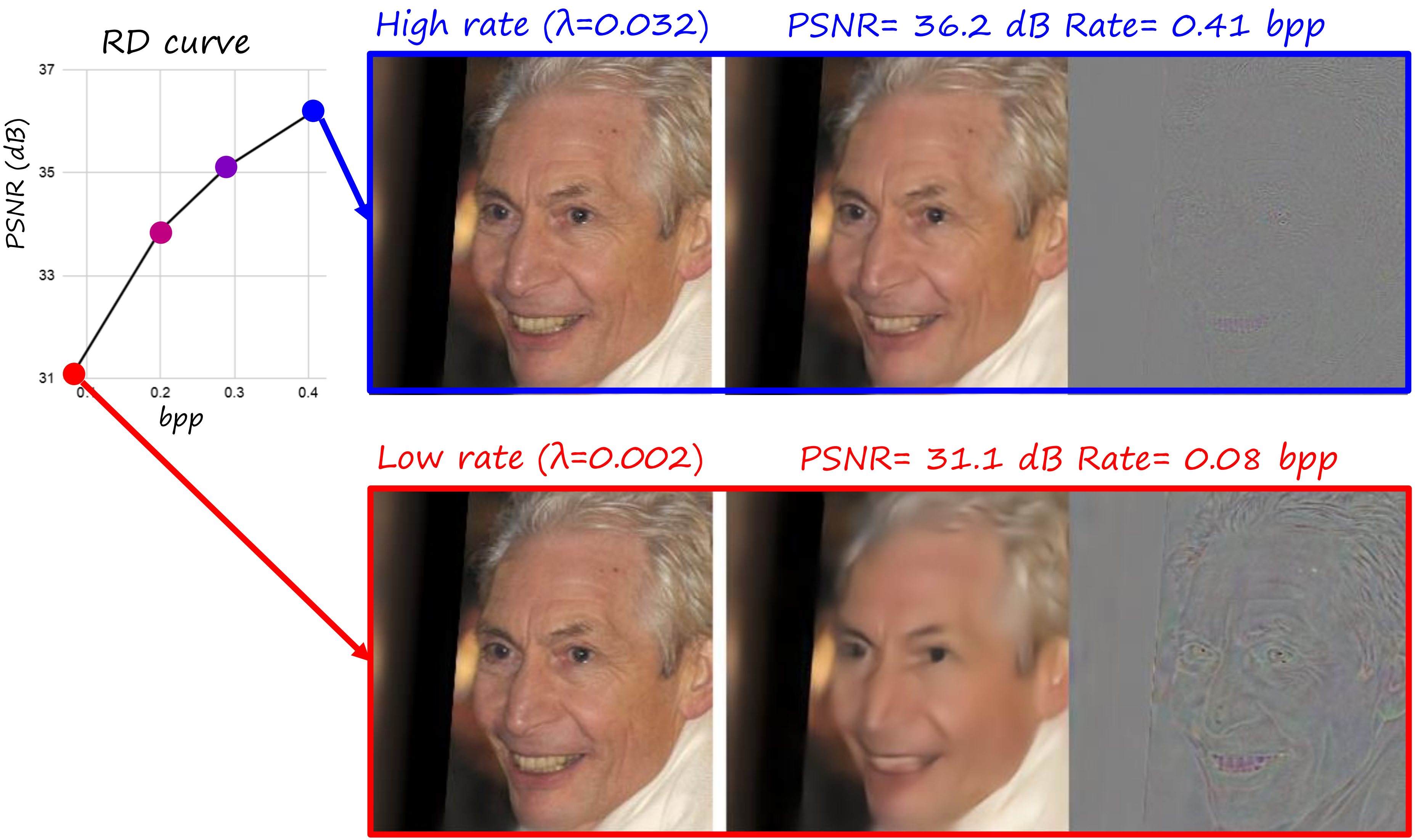
Generalized divisive normalization (GDN) [Balle2016]

$$\hat{y}_i = \frac{y_i}{\left(\beta_i + \sum_j \gamma_{ij}\, y_j^2\right)^{1/2}}$$



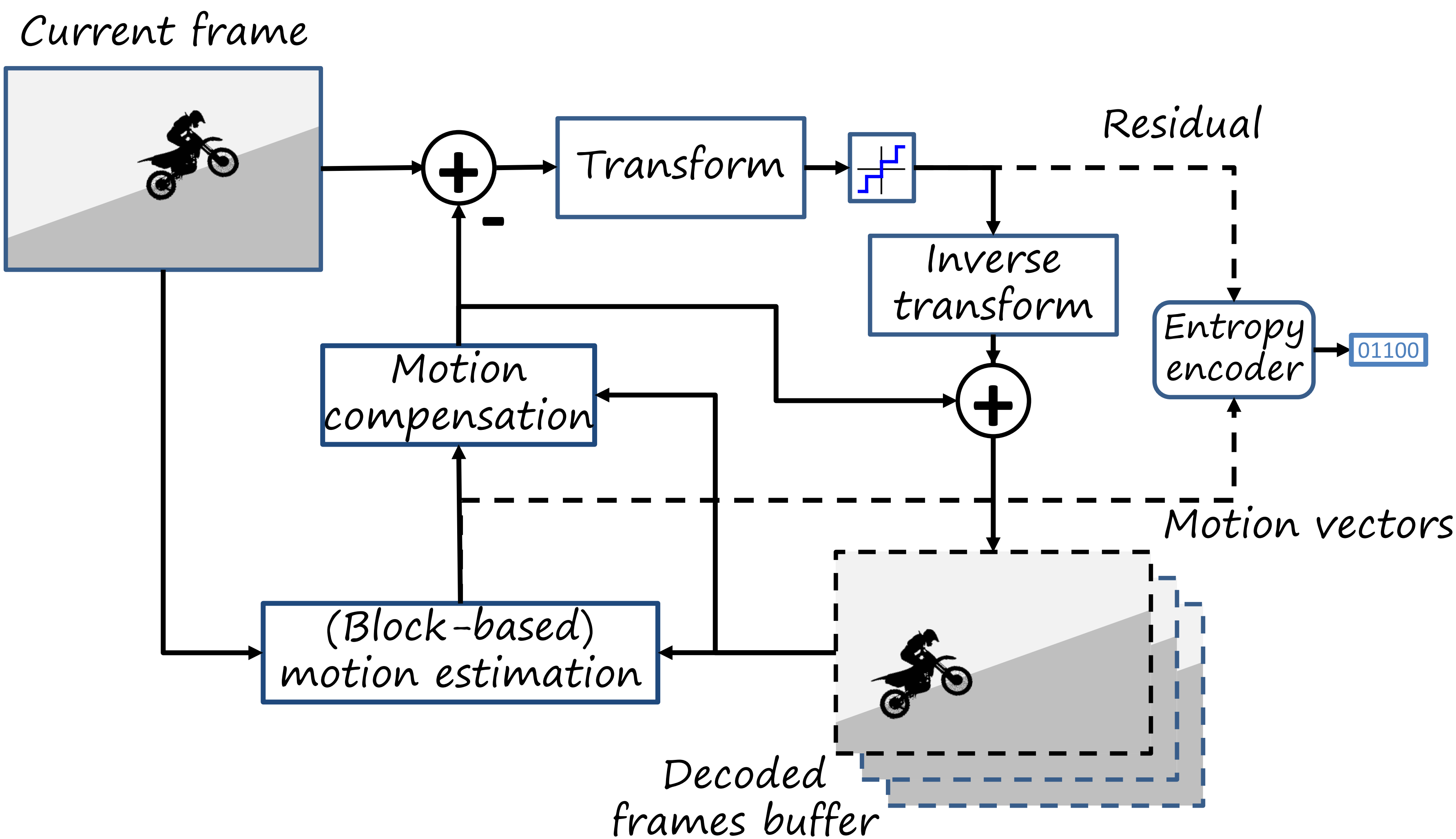
Learnable
parameters

Rate-distortion tradeoff λ



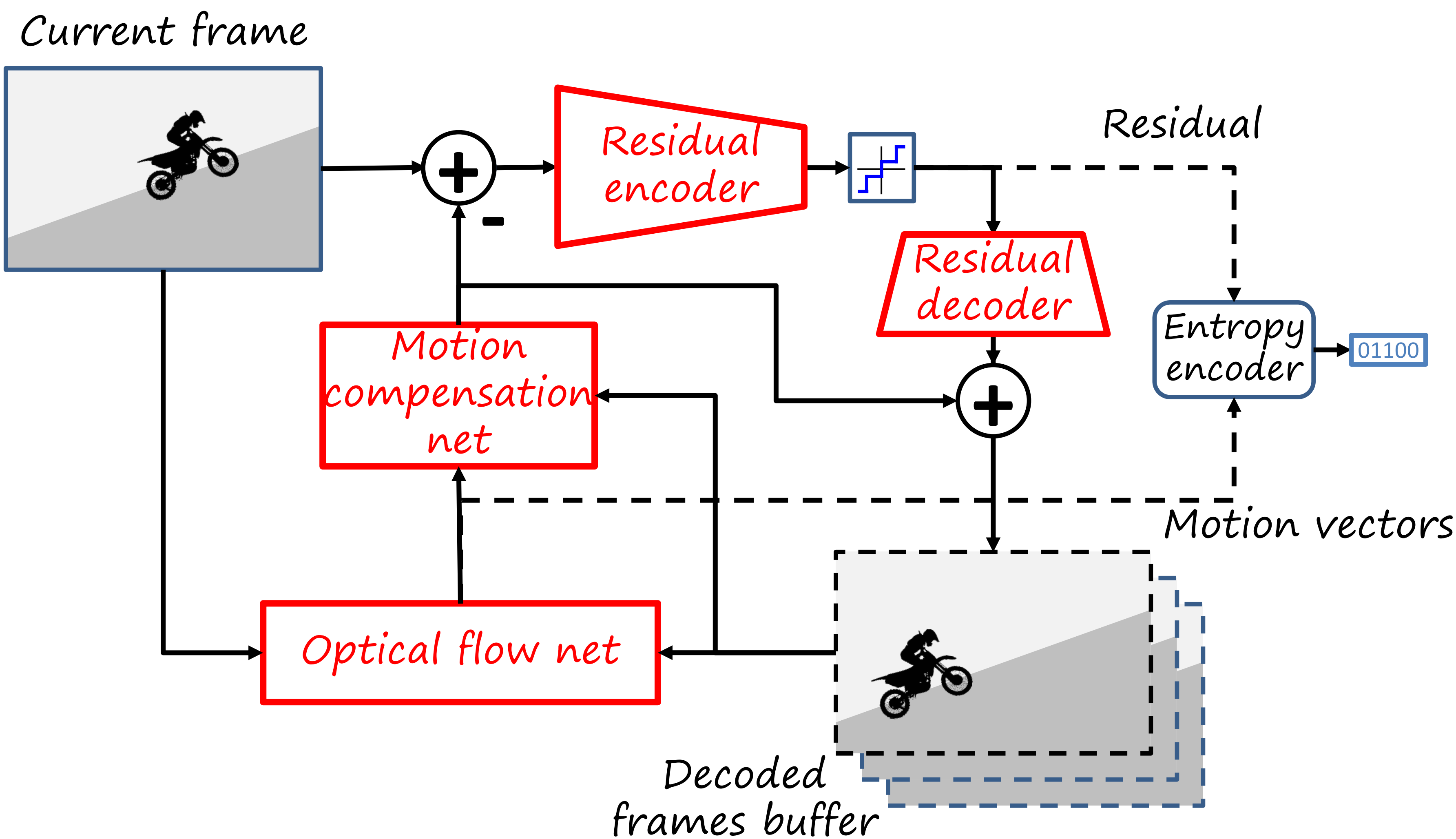
Traditional video compression

Replace modules by trainable neural networks



Neural video compression

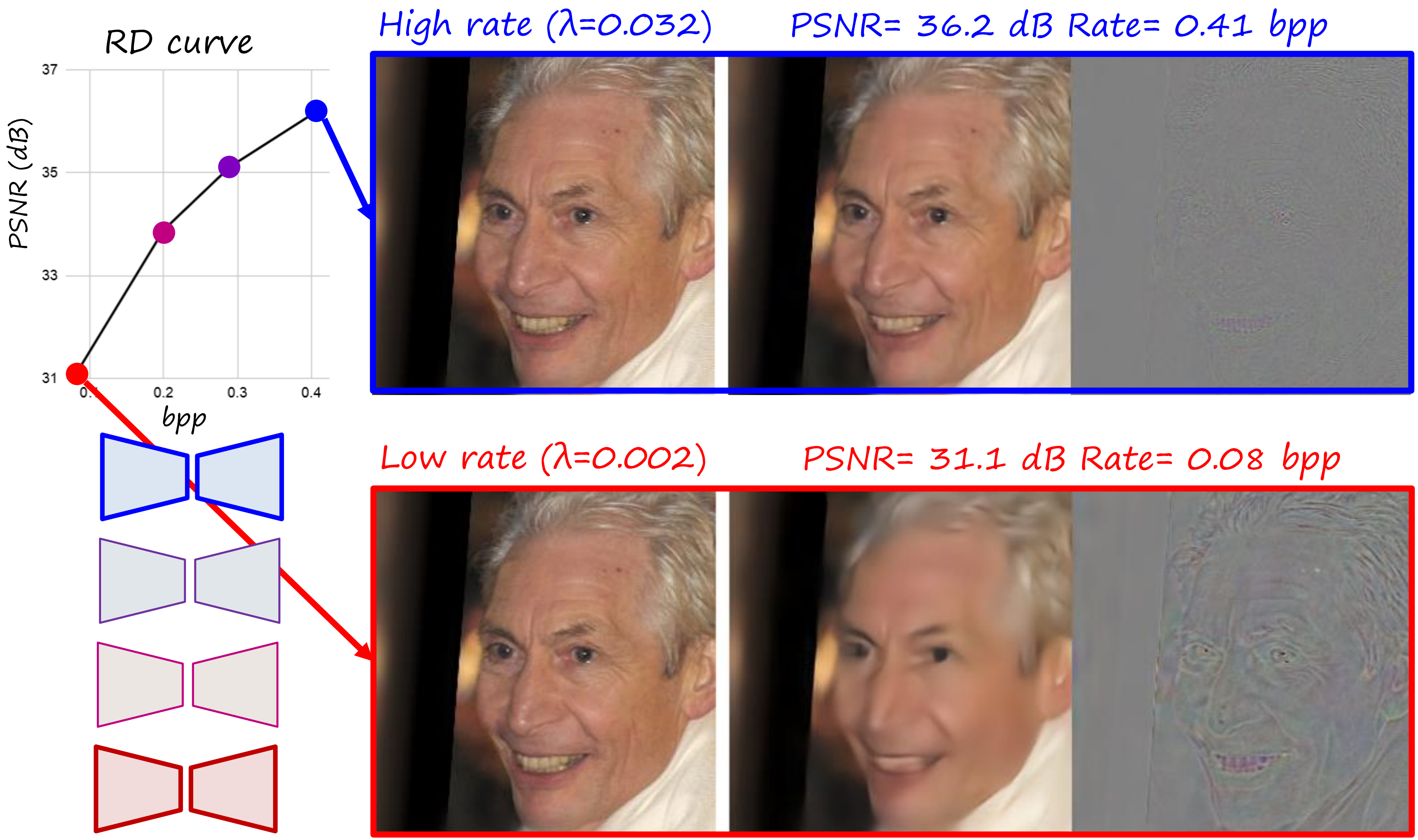
Replace modules by trainable neural networks



Outline

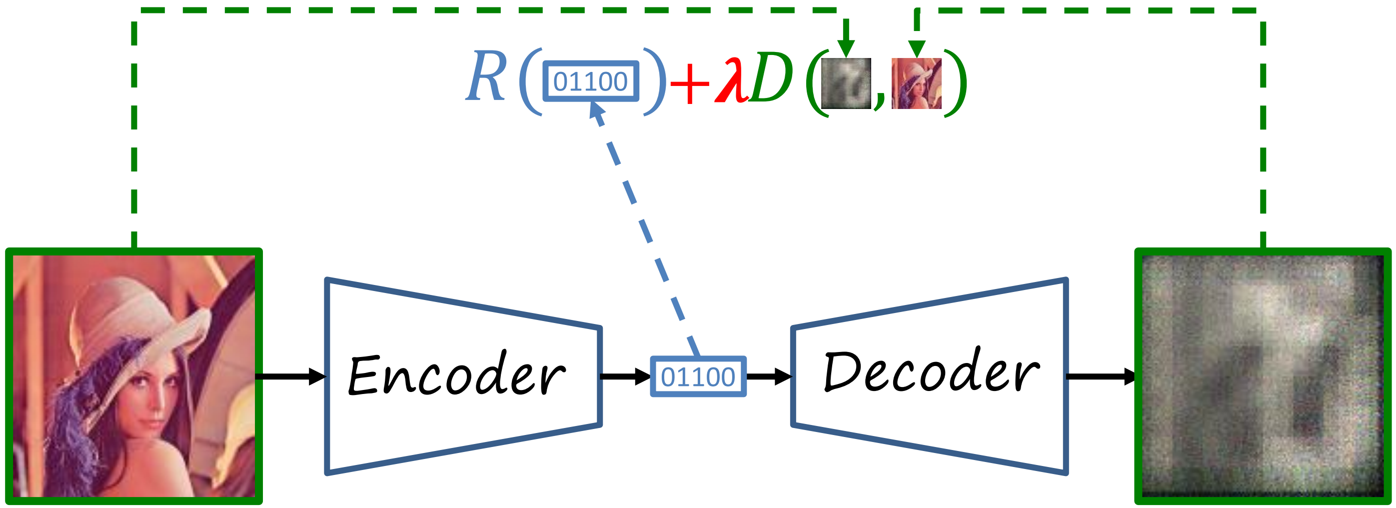
- Introduction: image/video coding
- Compression with neural networks
- Towards practical image compression
- Visual quality: perception vs distortion
- Video restoration and applications

Rate-distortion tradeoff λ



Problems: total memory, total training time

Is neural image compression practical?



Practical neural image compression?
- Minimize rate ✓
- Minimize distortion ✓
- Variable rate ✗
- Low memory ✗
- Low computation ✗
- Low latency ✗

MAE [SPL2020]

SlimCAE [CVPR2021]

Limitations
- λ is fixed
- Heavy encoders/decoders

Other practical considerations
- **Domain-specific codecs** (e.g. videoconference, screencast)
- **Back./forw. compatibility** (with legacy encoders/decoders)

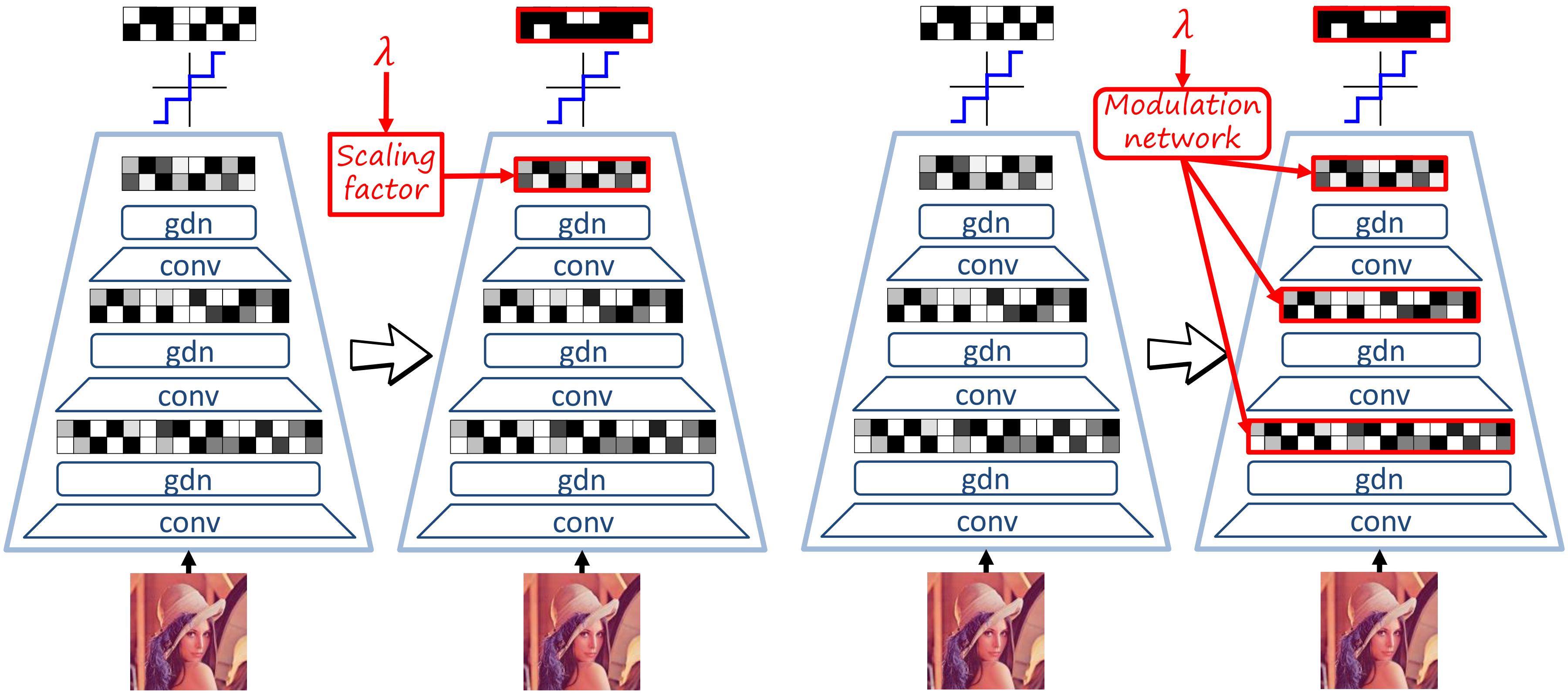
DANICE [CLIC2021]

[SPL2020] Variable Rate Deep Image Compression with Modulated Autoencoder, Signal Processing Letters 2020
[CVPR2021] Slimmable compressive autoencoders for practical imaga compression, CVPR 2021
[CLIC2021] DANICE: Domain adaptation without forgetting in neural image compression, CLIC 2021 at CVPR 2021

Variable rate with modulated autoencoders

Objective: one single model for multiple λ

Bottleneck scaling [Theis2017] | Feature modulation [MAE, cAE]



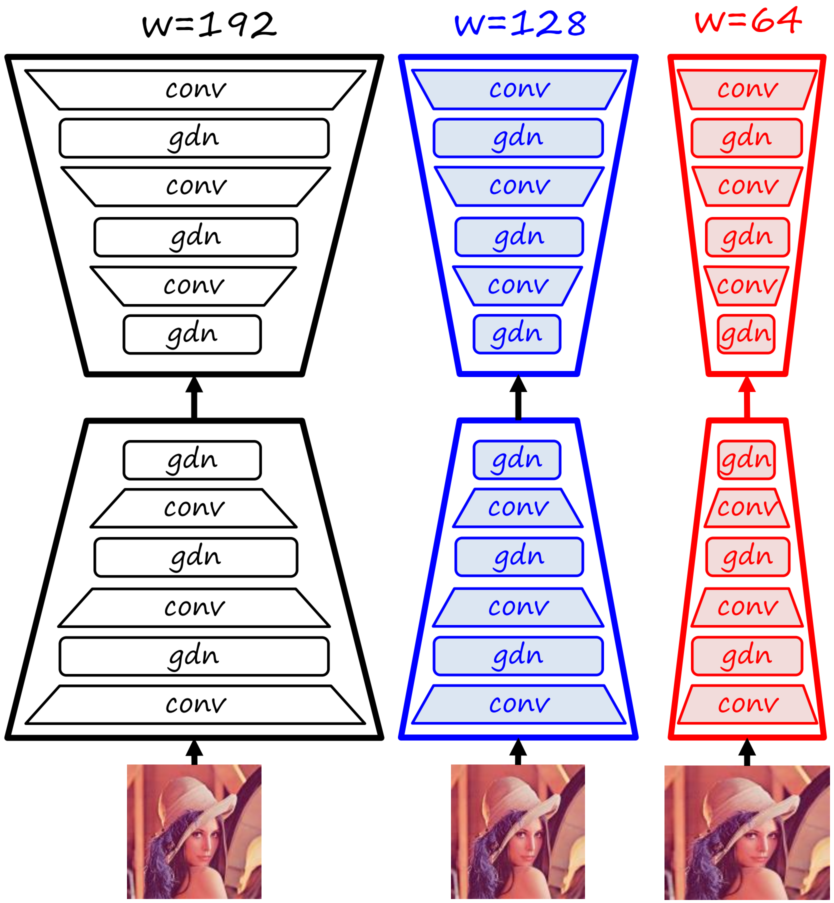
- Minimize rate ✓
- Minimize distortion ✓
- Variable rate ✓

- Low memory ✗
- Low computation ✗
- Low latency ✗

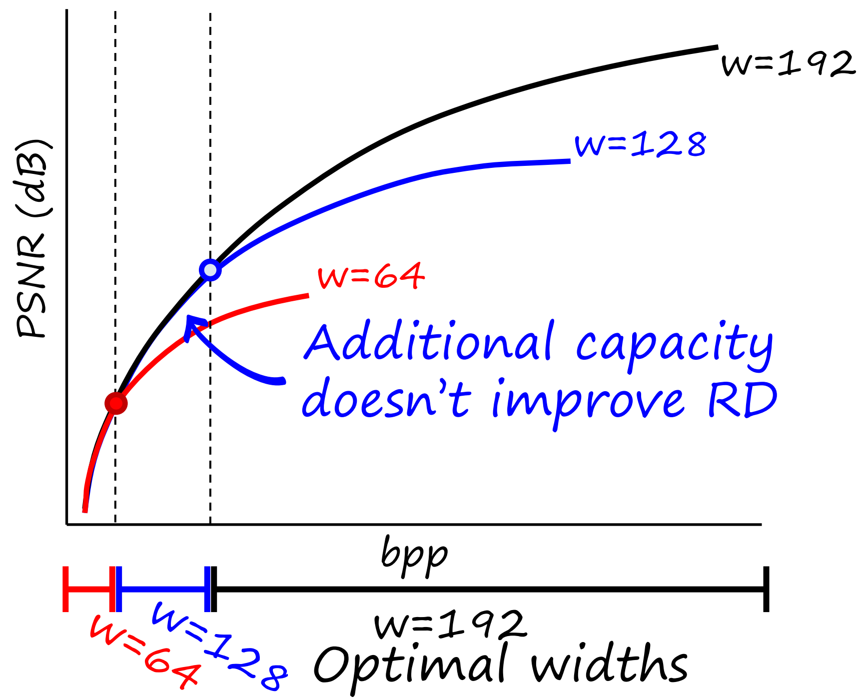
cAE: conditional autoencoder [Choi2019]
MAE: modulated autoencoder [Yang2020]

Model capacity and rate-distortion

w=filters per layer



There is a **minimal capacity** for every RD tradeoff



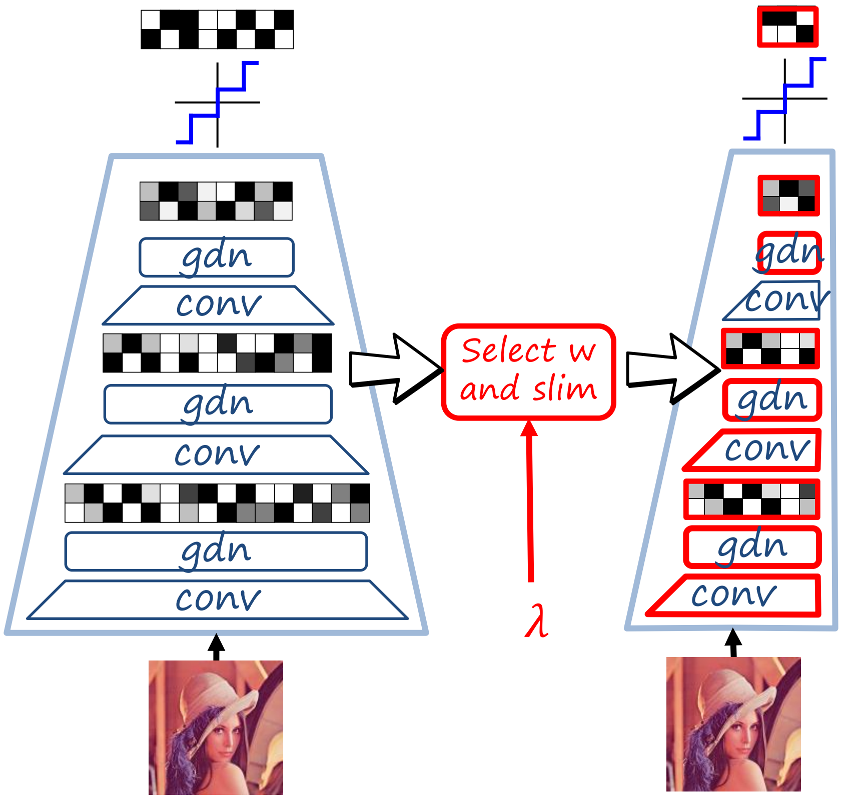
Lower w results in less memory and computation!!

[Slimmable compressive autoencoders for practical imaga compression](), CVPR 2021

Slimmable compressive autoencoder

Approach: slim the network to the minimal capacity for a given λ

Slimming [SlimCAE]

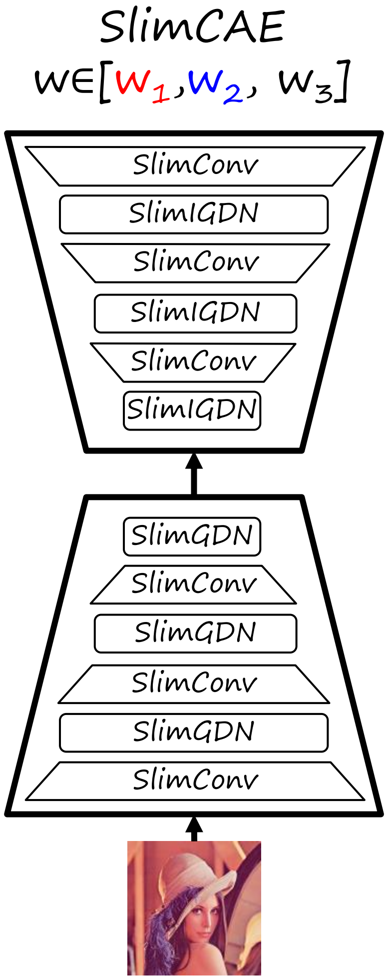


- Minimize rate ✓
- Minimize distortion ✓
- Variable rate ✓
- Lower memory ✓
- Lower computation ✓
- Lower latency ✓

(for low-mid rates)

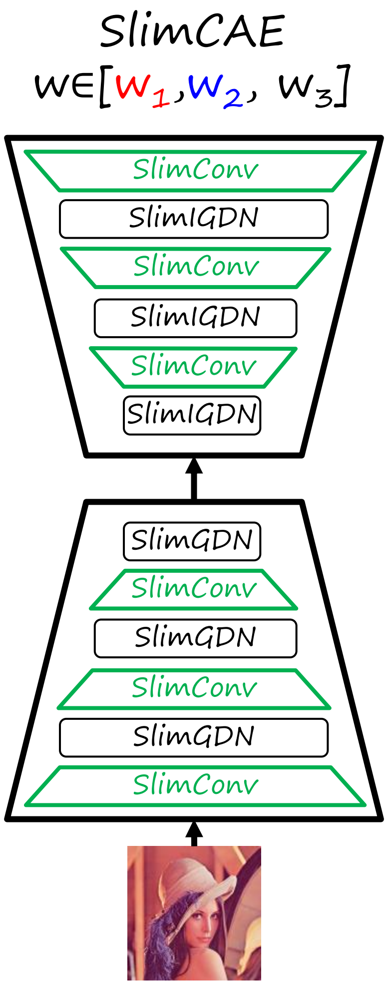
Slimmable compressive autoencoders for practical imaga compression, CVPR 2021

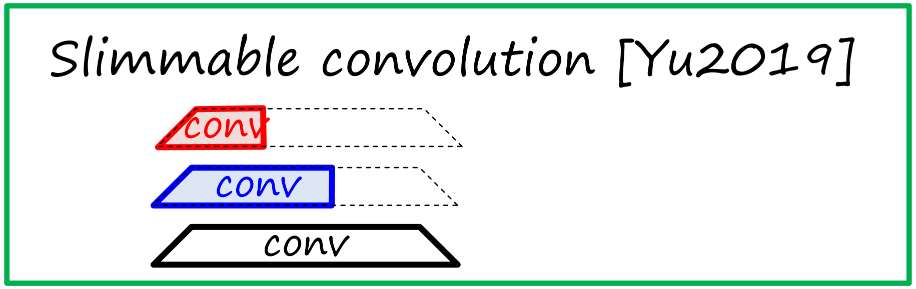
Slimmable layers in SlimCAE



Slimmable compressive autoencoders for practical imaga compression, CVPR 2021

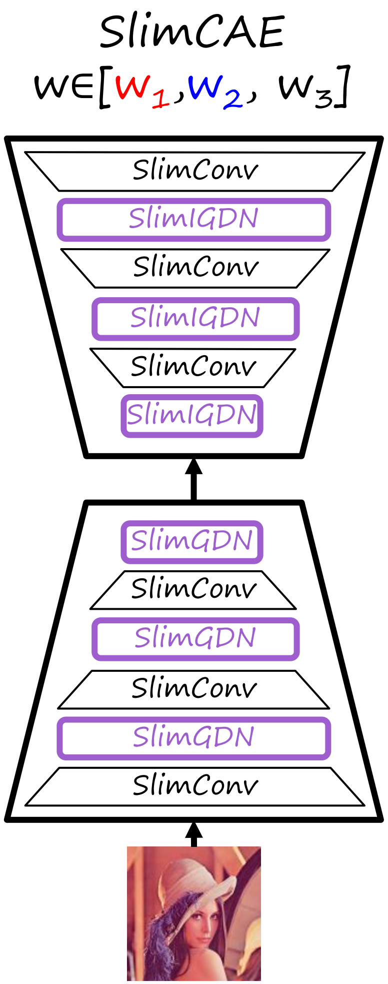
Slimmable layers in SlimCAE

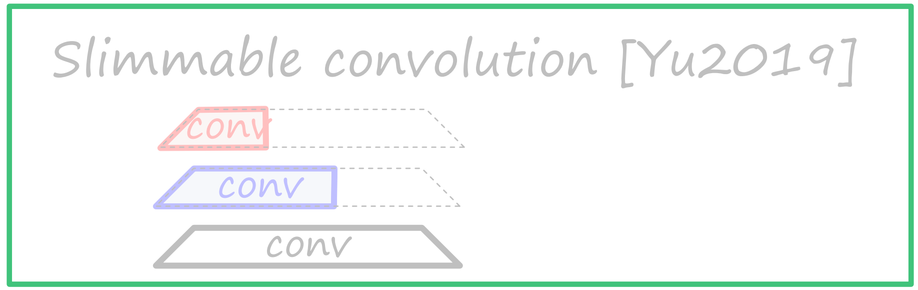


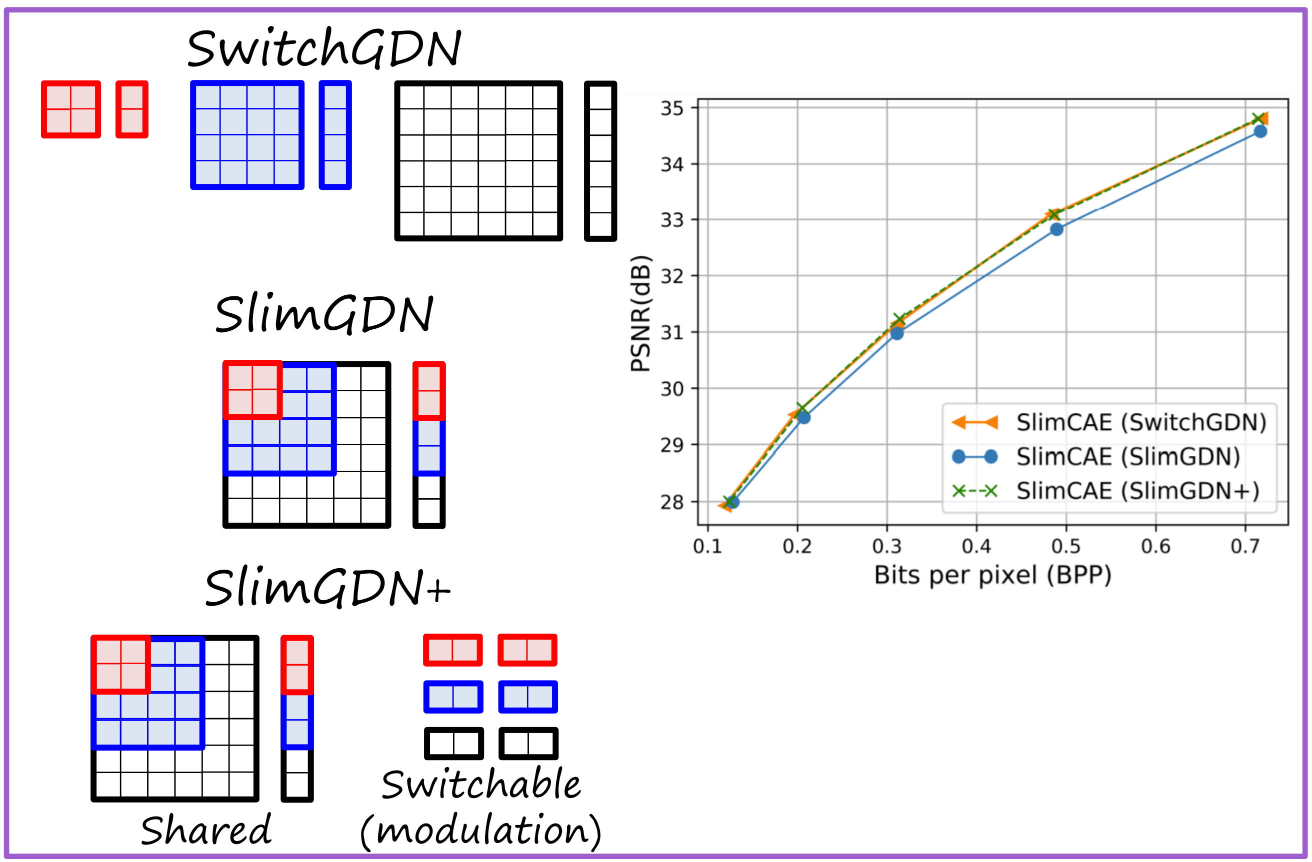


Slimmable compressive autoencoders for practical imaga compression, CVPR 2021

Slimmable layers in SlimCAE





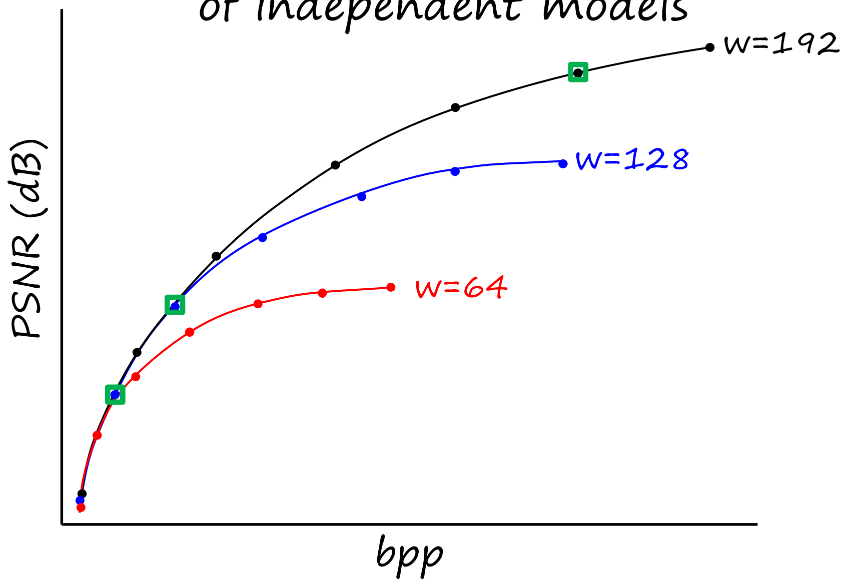


Slimmable compressive autoencoders for practical imaga compression, CVPR 2021

Training SlimCAE

Problem: we need the ***optimal* λs** to train the SlimCAE

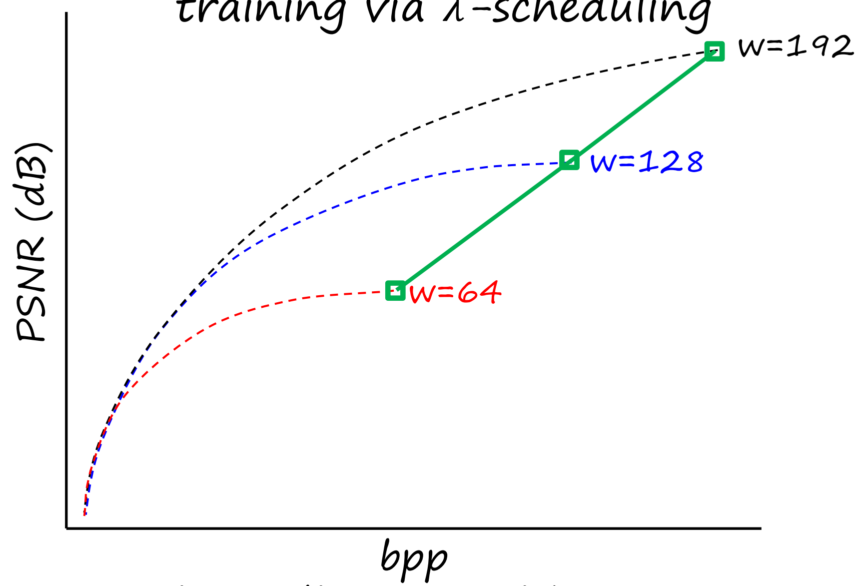
Estimate from RD curves of independent models



1. Train several independent models for different w
2. Plot RD curves and find critical points
3. Estimate optimal λs from trained models

Problem: extremely expensive!

Automatically estimate during training via λ-scheduling



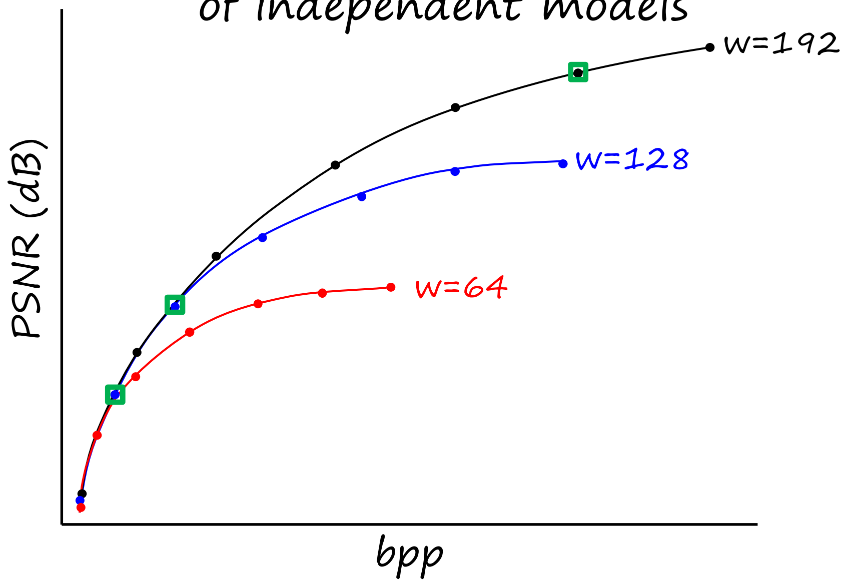
1. Train a SlimCAE with $\lambda_1 = \lambda_2 = \lambda_3$
2. While not converged do
   - Update λs according to schedule
   - Optimize CAE

Slimmable compressive autoencoders for practical imaga compression, CVPR 2021

Training SlimCAE

Problem: we need the optimal λs to train the SlimCAE

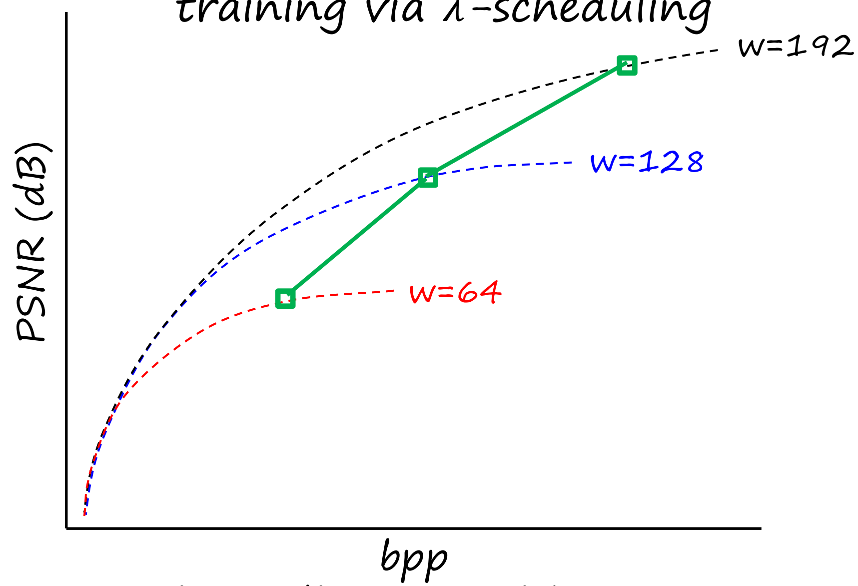
Estimate from RD curves of independent models



1. Train several independent models for different w
2. Plot RD curves and find critical points
3. Estimate optimal λs from trained models

Problem: extremely expensive!

Automatically estimate during training via λ-scheduling



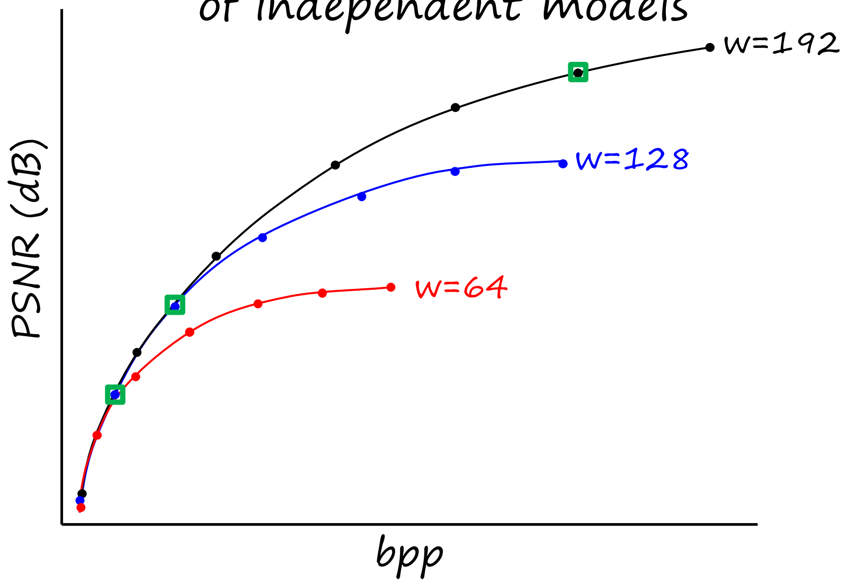
1. Train a SlimCAE with $\lambda_1 = \lambda_2 = \lambda_3$
2. While not converged do
   - Update λs according to schedule
   - Optimize CAE

Slimmable compressive autoencoders for practical imaga compression, CVPR 2021

Training SlimCAE

Problem: we need the optimal λs to train the SlimCAE

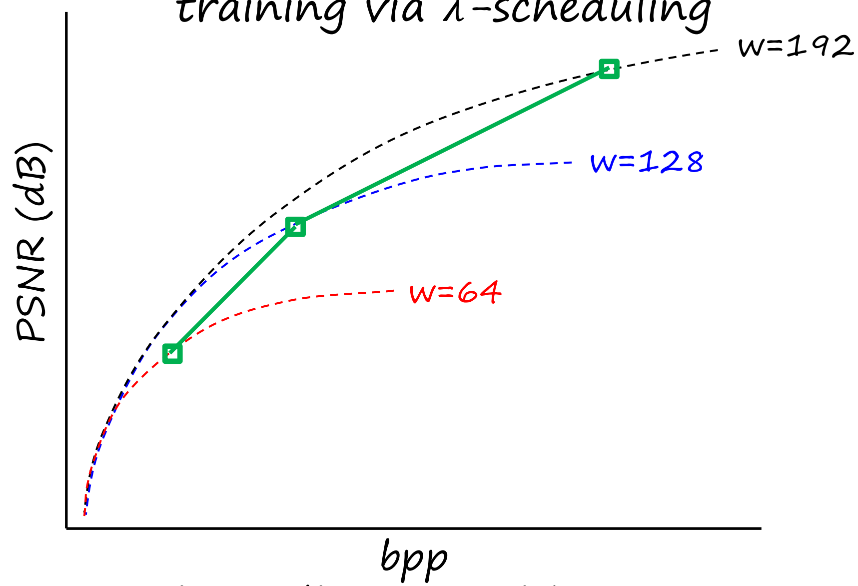
Estimate from RD curves of independent models



1. Train several independent models for different w
2. Plot RD curves and find critical points
3. Estimate optimal λs from trained models

Problem: extremely expensive!

Automatically estimate during training via λ-scheduling



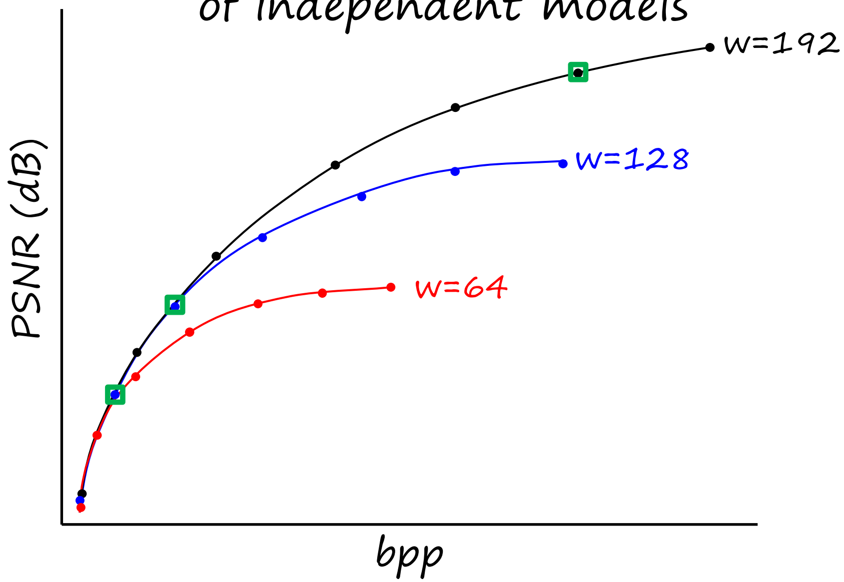
1. Train a SlimCAE with $\lambda_1 = \lambda_2 = \lambda_3$
2. While not converged do
   - Update λs according to schedule
   - Optimize CAE

Slimmable compressive autoencoders for practical imaga compression, CVPR 2021

Training SlimCAE

Problem: we need the optimal λs to train the SlimCAE

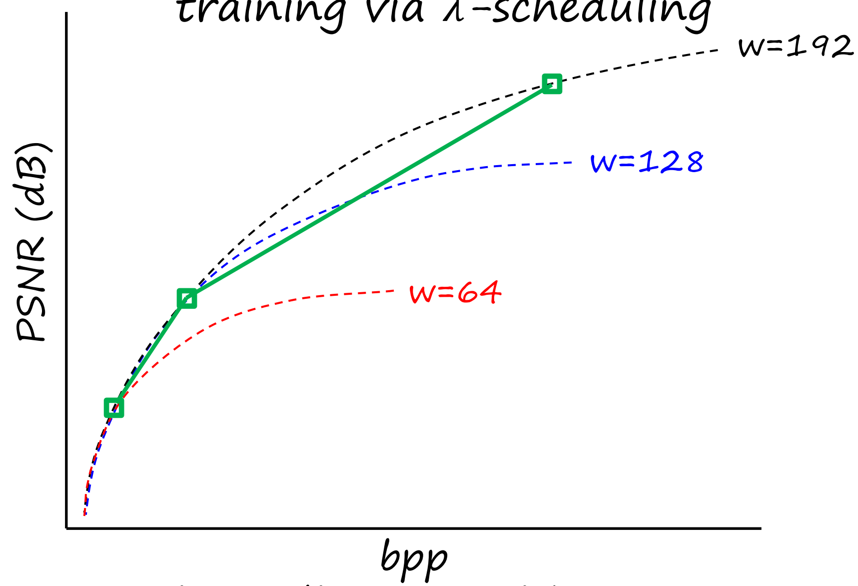
Estimate from RD curves of independent models



1. Train several independent models for different w
2. Plot RD curves and find critical points
3. Estimate optimal λs from trained models

Problem: extremely expensive!

Automatically estimate during training via λ-scheduling

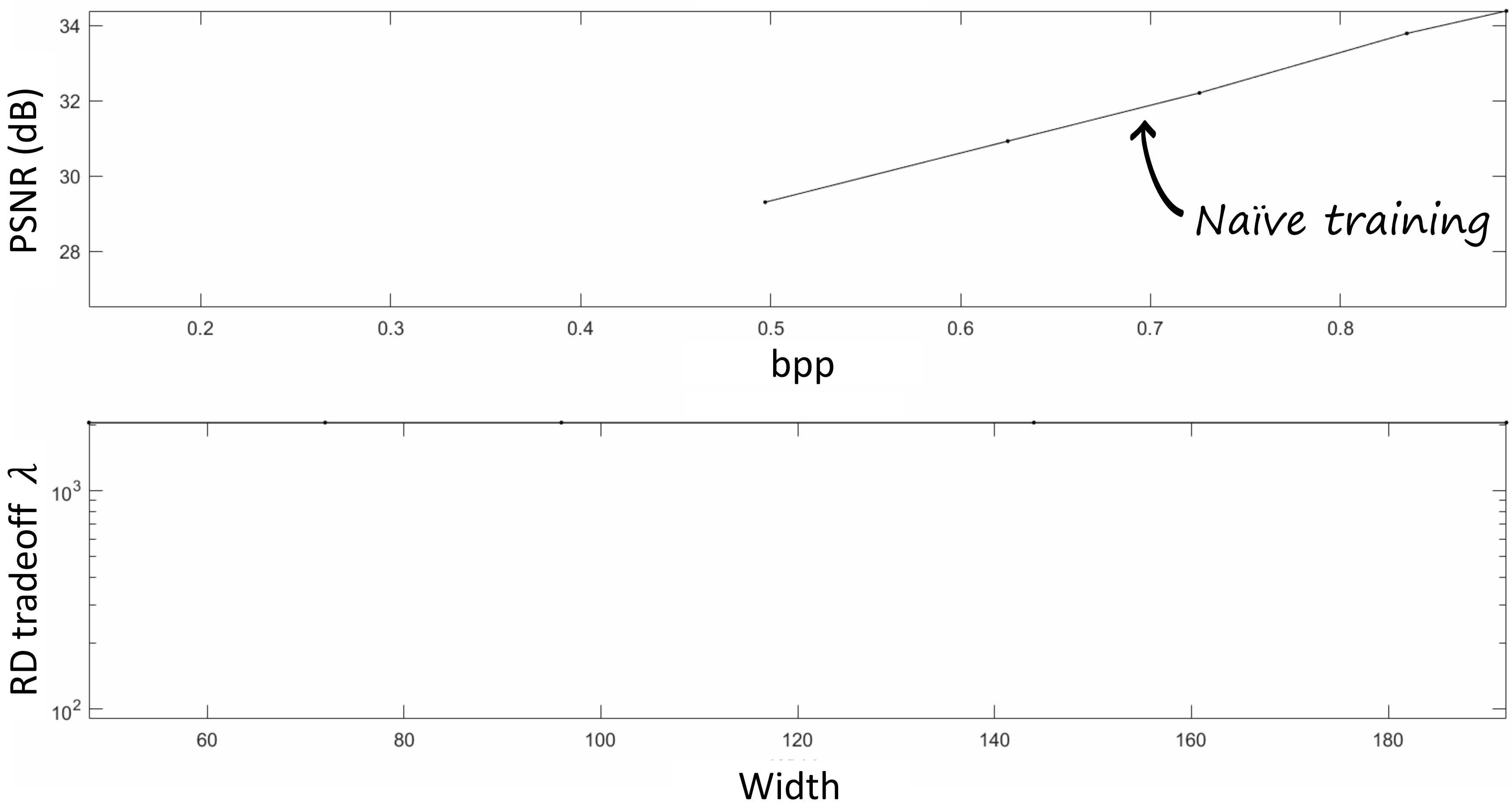


1. Train a SlimCAE with $\lambda_1 = \lambda_2 = \lambda_3$
2. While not converged do
   - Update λs according to schedule
   - Optimize CAE

Directly train one model!

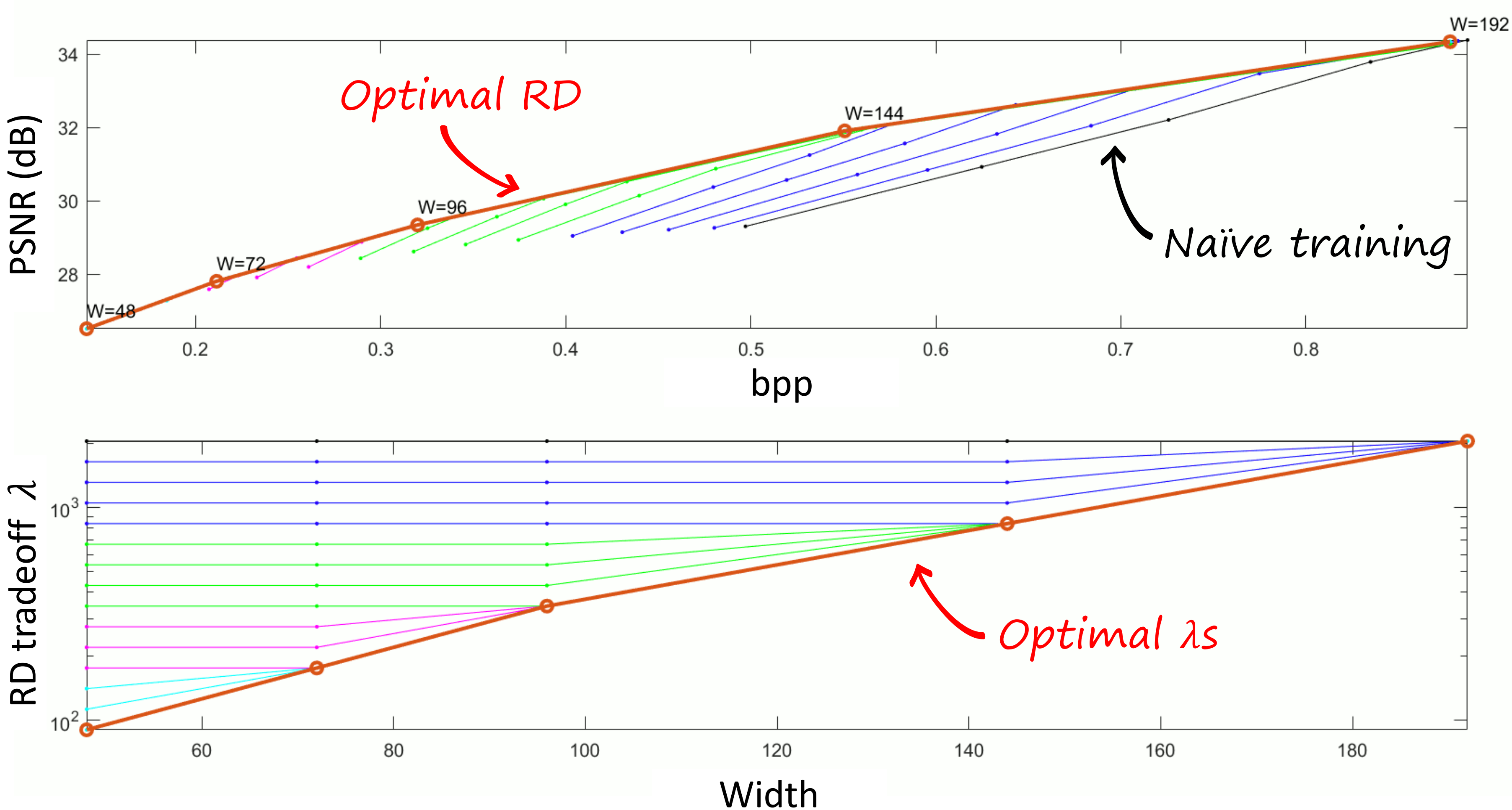
Slimmable compressive autoencoders for practical imaga compression, CVPR 2021

λ-scheduling. Example

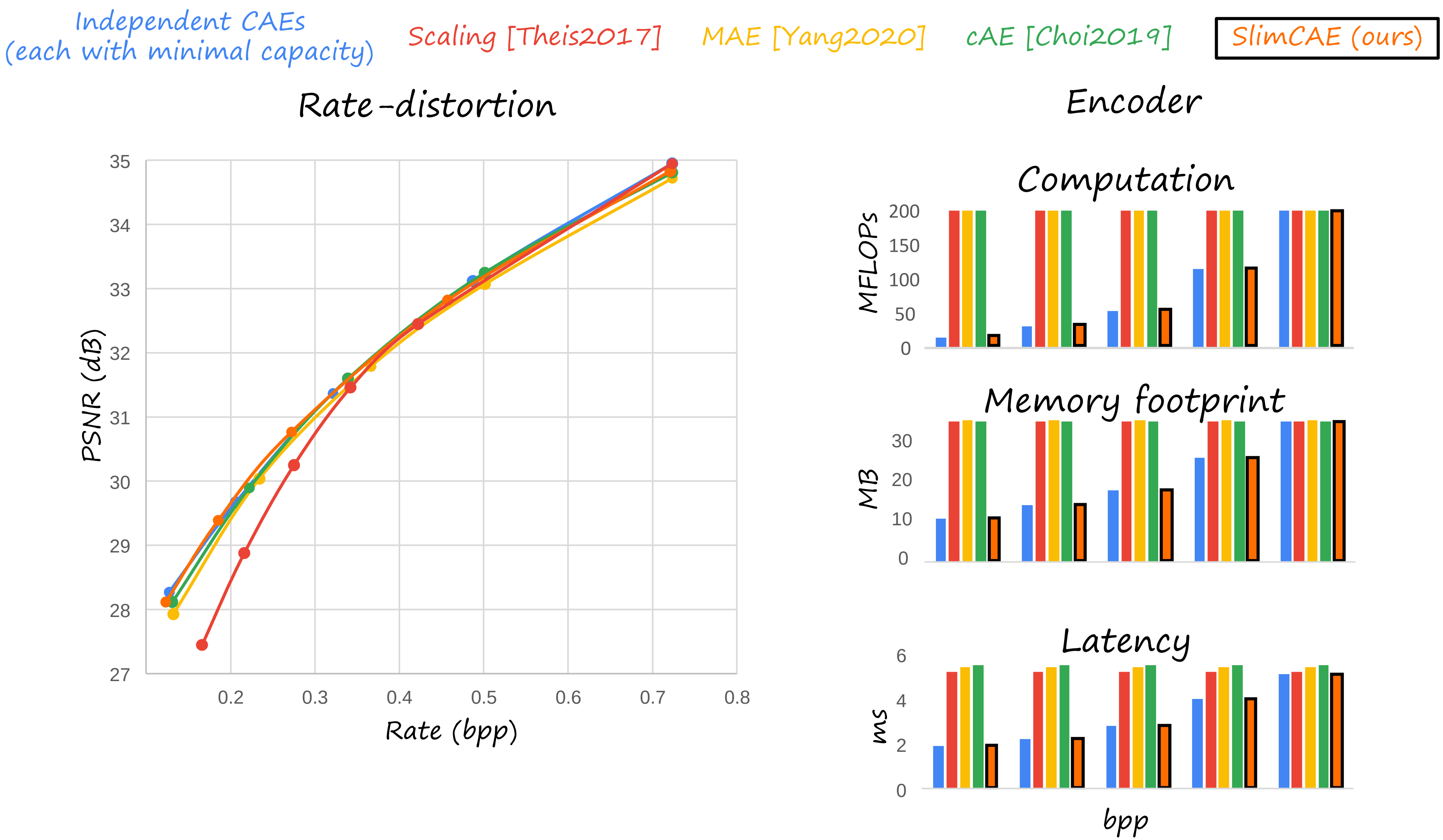


[Slimmable compressive autoencoders for practical imaga compression](), CVPR 2021

λ-scheduling

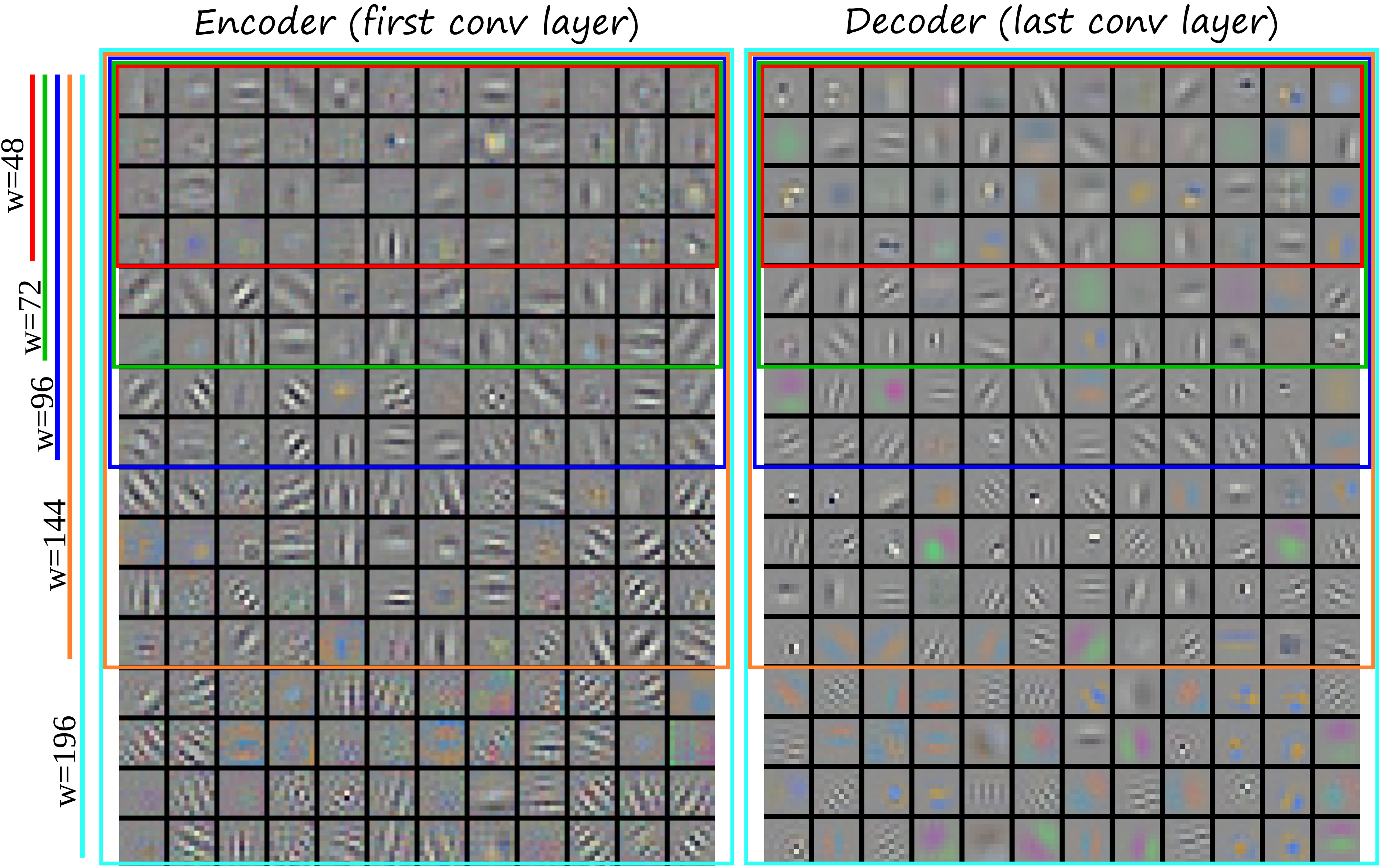


Performance comparison



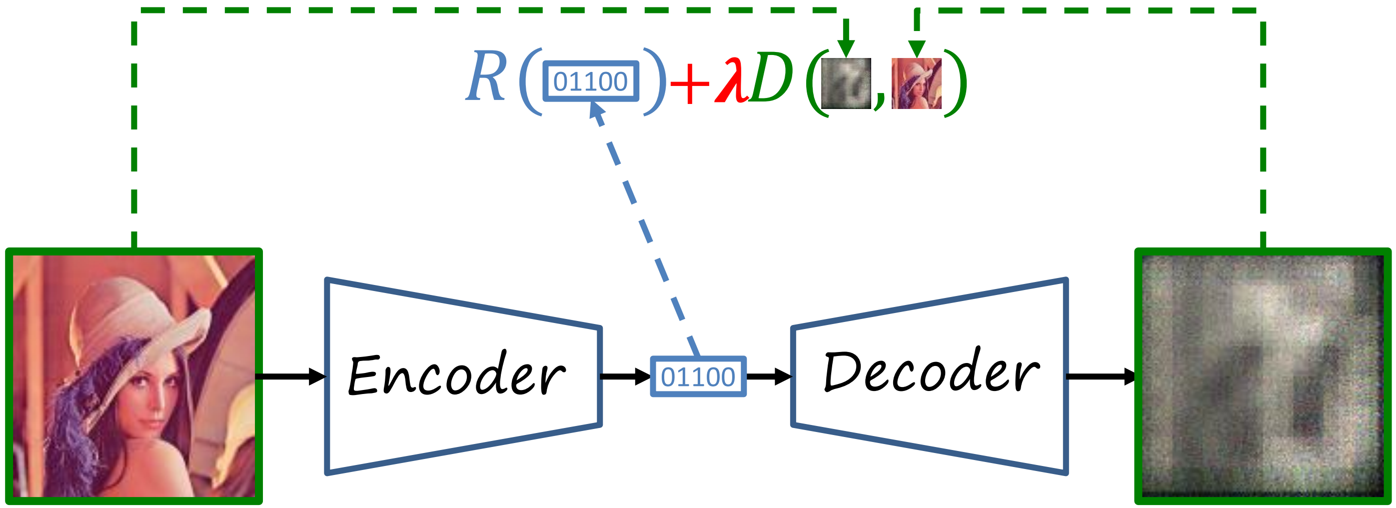
[Slimmable compressive autoencoders for practical imaga compression](), CVPR 2021

Visualizing some parameters



[Slimmable compressive autoencoders for practical imaga compression](), CVPR 2021

Is neural image compression practical?



Practical neural image compression?
- Minimize rate ✓
- Minimize distortion ✓
- Variable rate ✗
- Low memory ✗
- Low computation ✗
- Low latency ✗

MAE [SPL2020]

SlimCAE [CVPR2021]

Limitations
- λ is fixed
- Heavy encoders/decoders

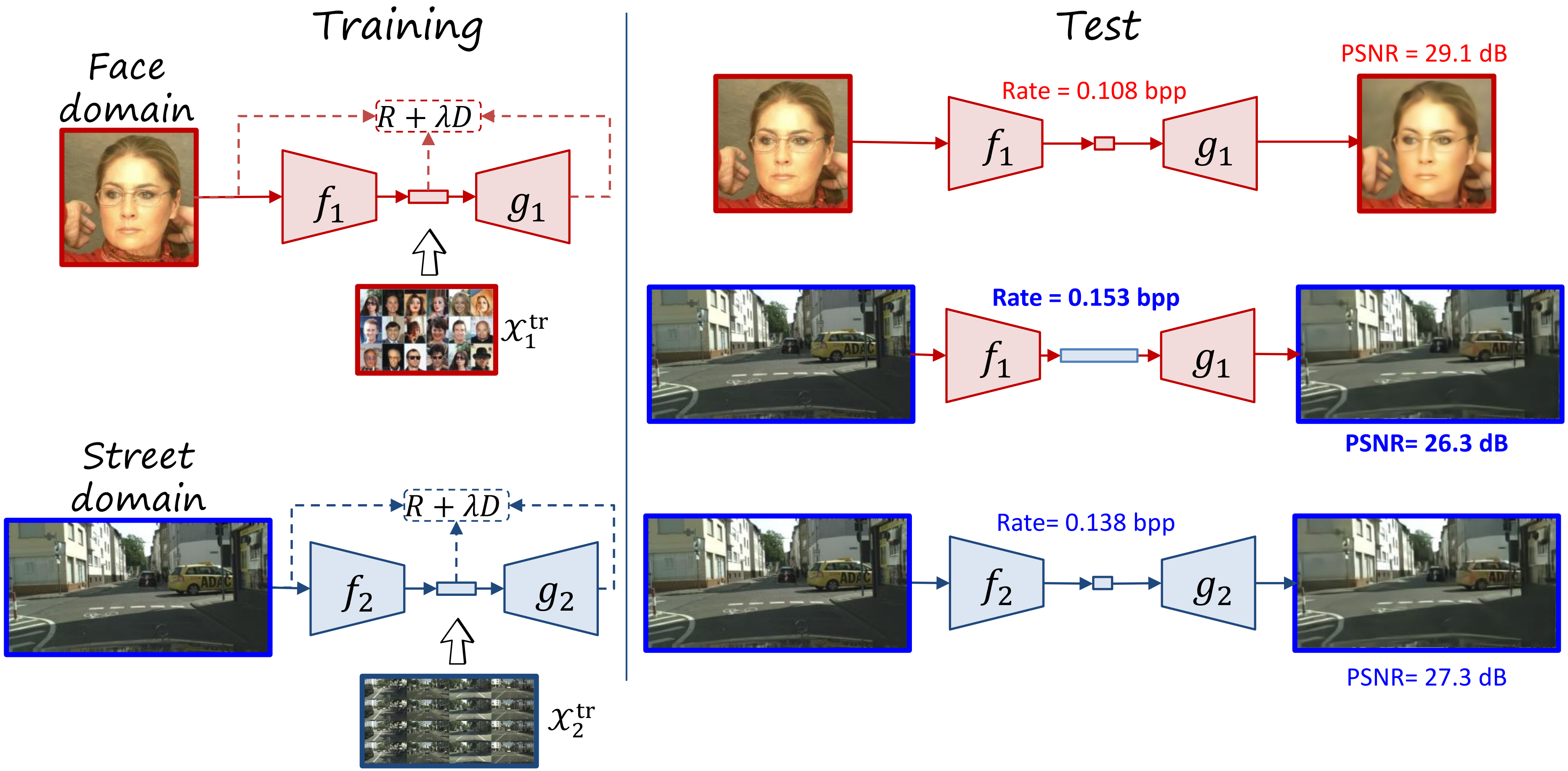
Other practical considerations
- **Domain-specific codecs** (e.g. videoconference, screencast)
- **Back./forw. compatibility** (with legacy encoders/decoders)

DANICE [CLIC2021]

[SPL2020] Variable Rate Deep Image Compression with Modulated Autoencoder, Signal Processing Letters 2020
[CVPR2021] Slimmable compressive autoencoders for practical imaga compression, CVPR 2021
[CLIC2021] DANICE: Domain adaptation without forgetting in neural image compression, CLIC 2021 at CVPR 2021

Rate-distortion optimality of learned codecs

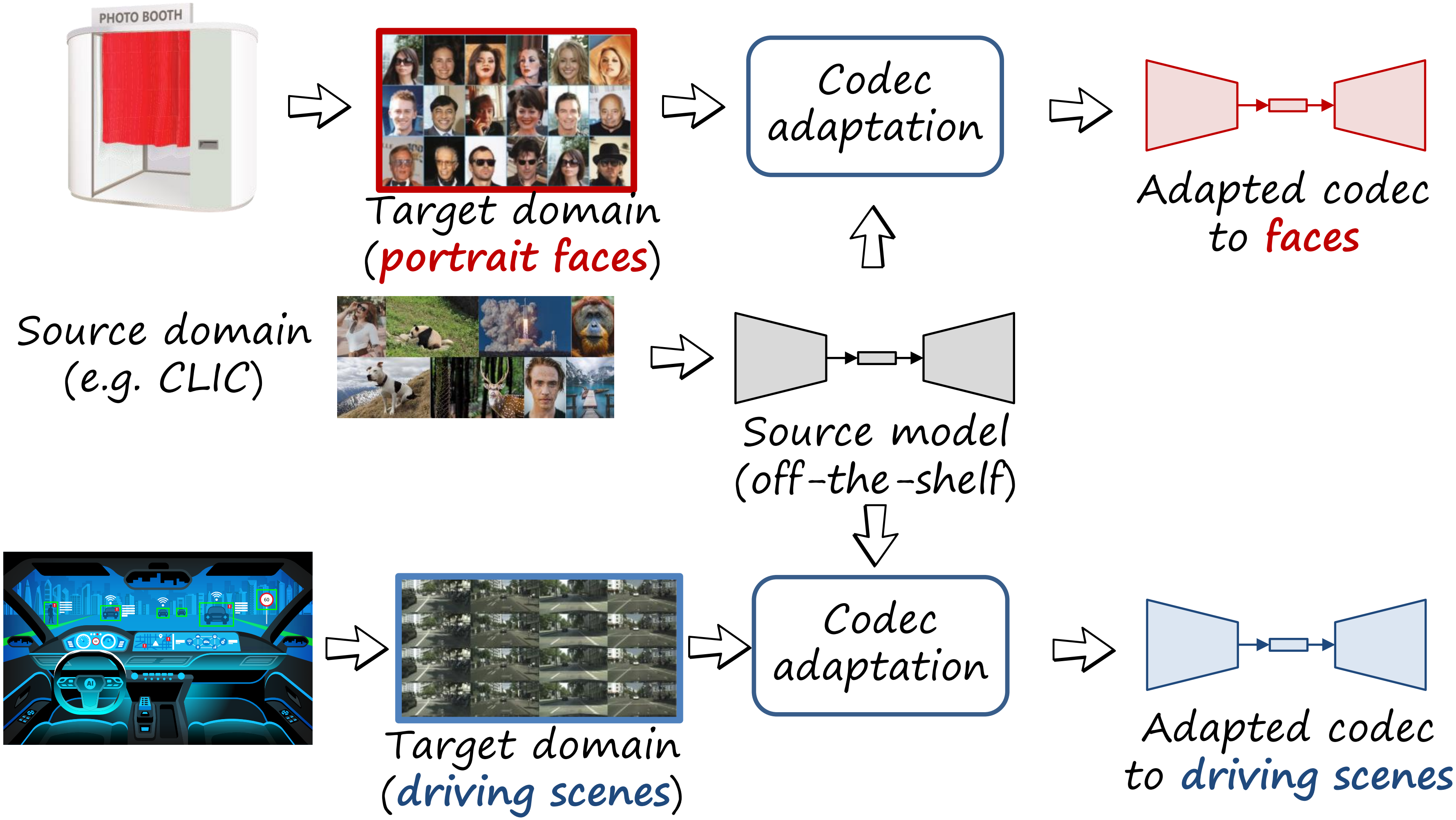
Learned codecs are only optimal in the domain of the training data



DANICE: Domain adaptation without forgetting in neural image compression, CLIC 2021 at CVPR 2021

Domain Adaptation in Neural Image ComprEssion (DANICE)

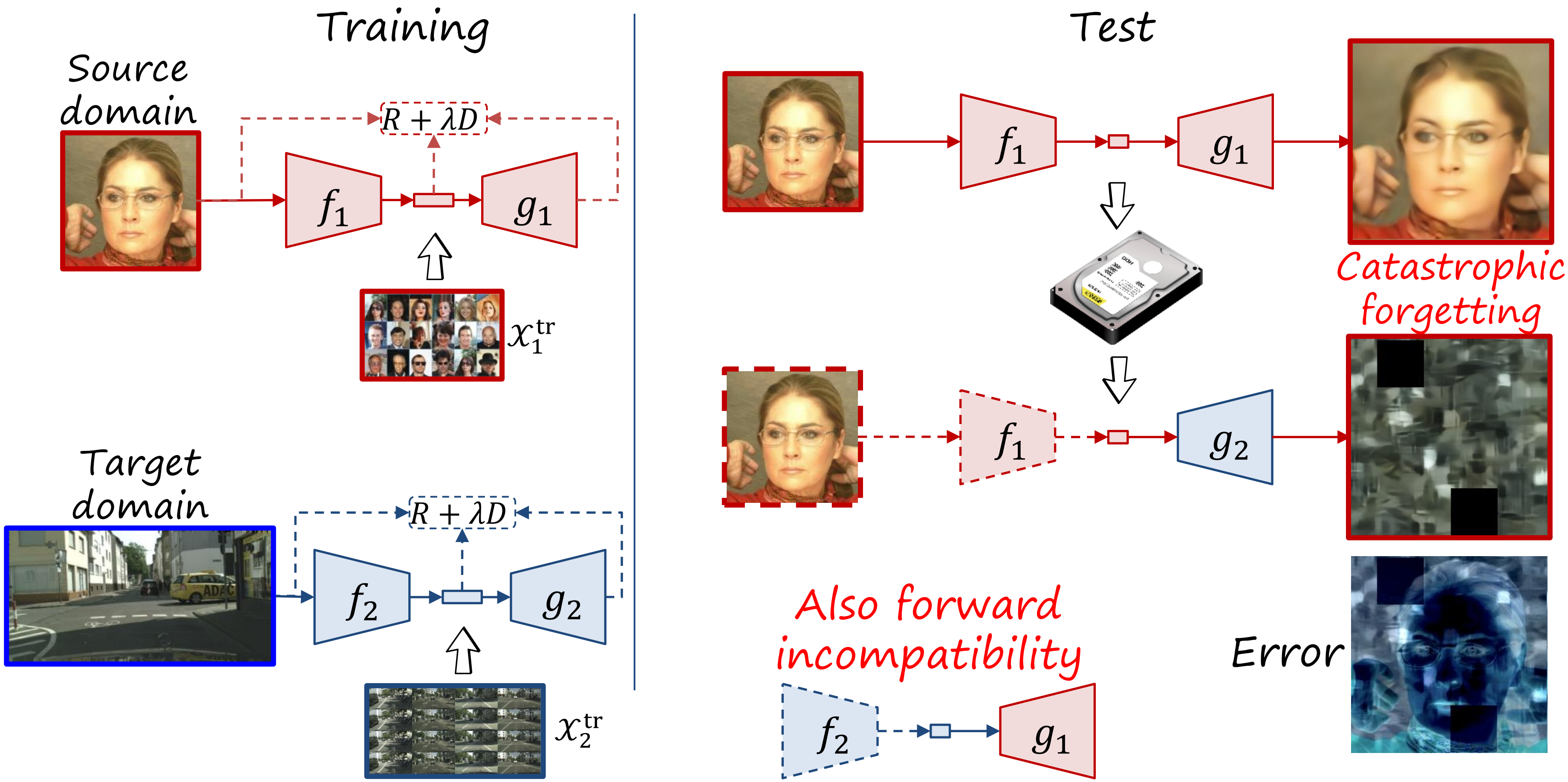
Learned codecs can be **customized with user content** to specific domains
Problem: usually **not enough custom data**; training is **expensive**
Solution: **transfer pre-trained codecs**



DANICE: Domain adaptation without forgetting in neural image compression, CLIC 2021 at CVPR 2021

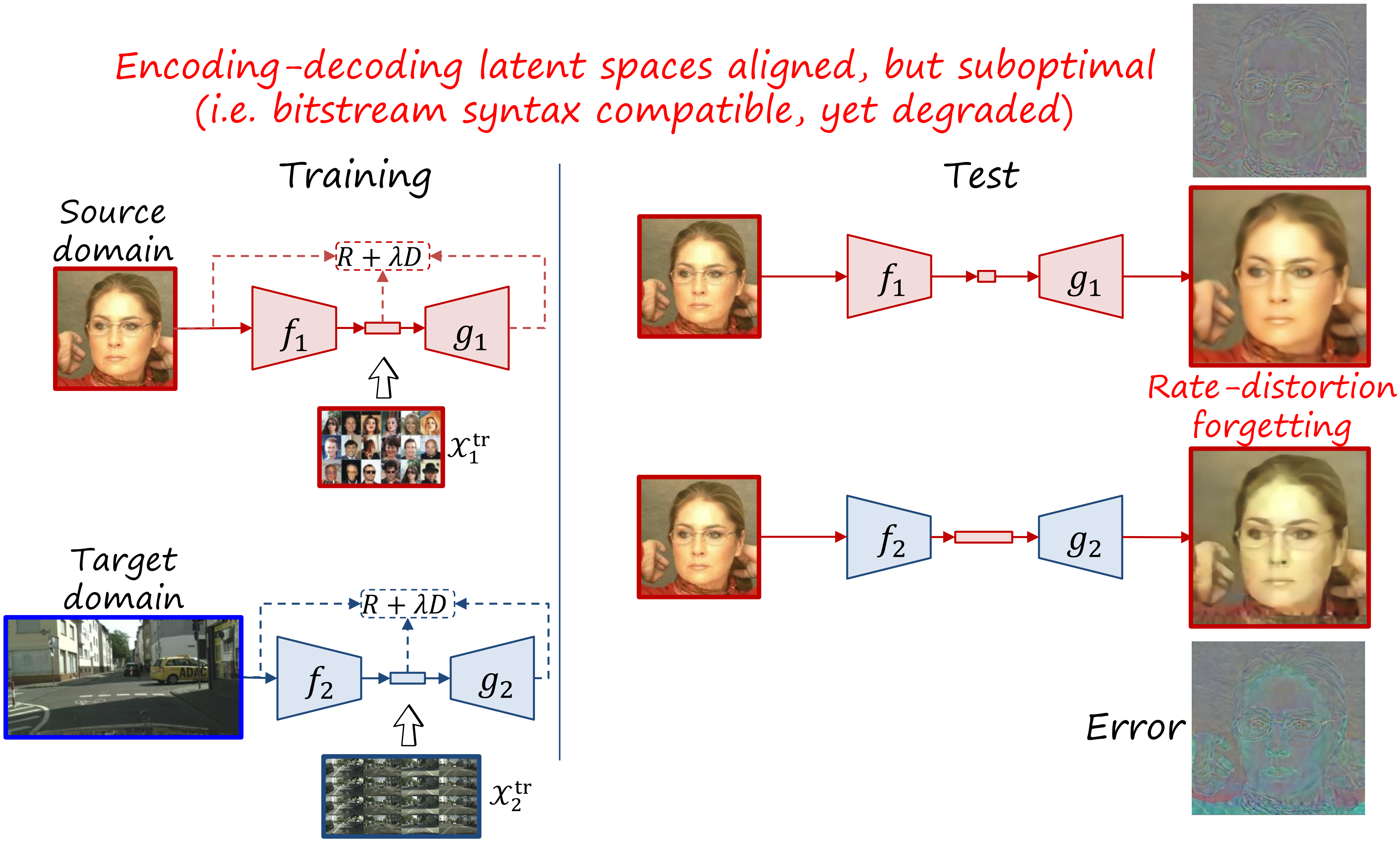
Backward incompatibility with legacy bitstreams: catastrophic forgetting

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



DANICE: Domain adaptation without forgetting in neural image compression, CLIC 2021 at CVPR 2021

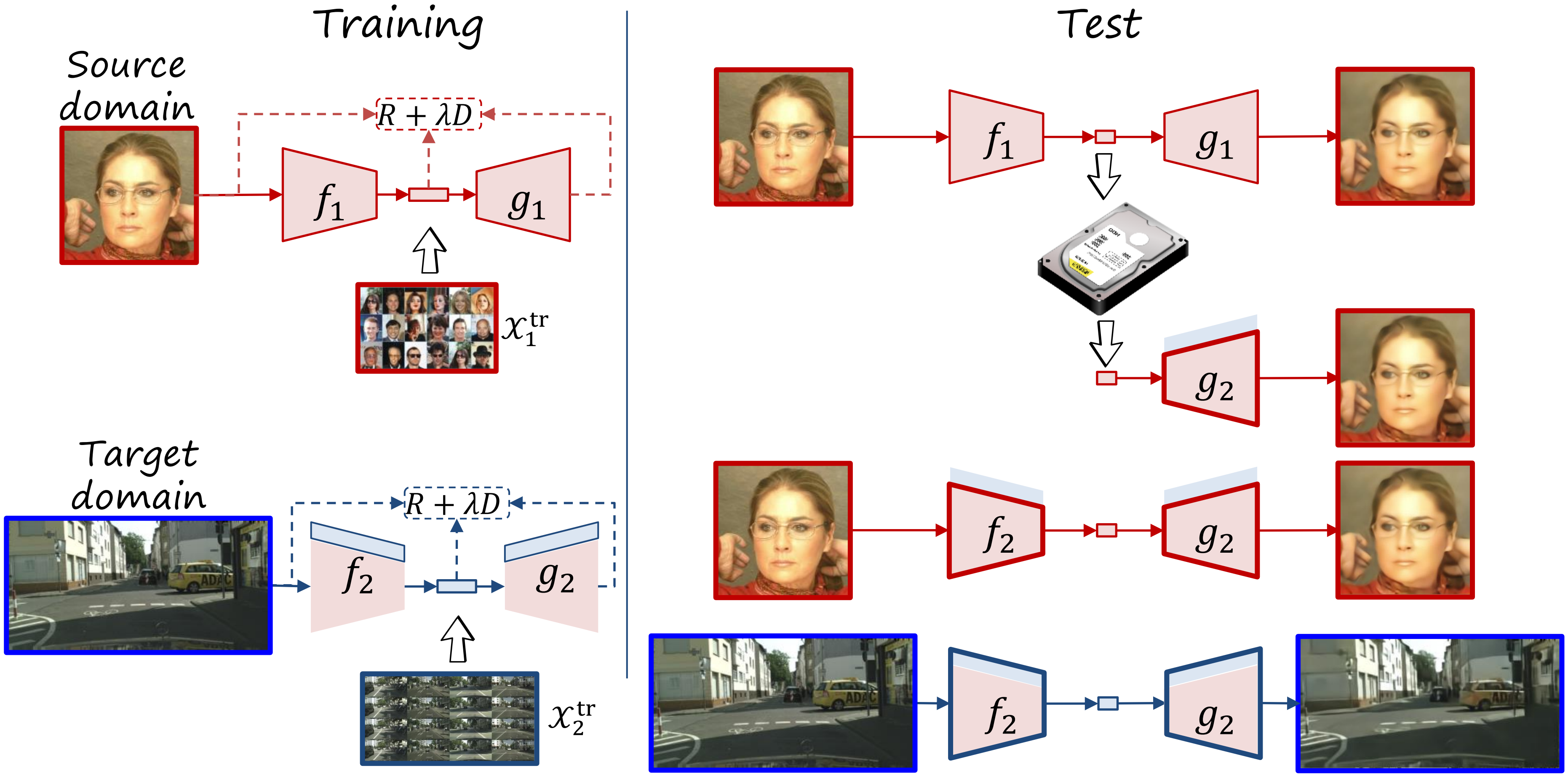
Rate-distortion forgetting



DANICE: Domain adaptation without forgetting in neural image compression, CLIC 2021 at CVPR 2021

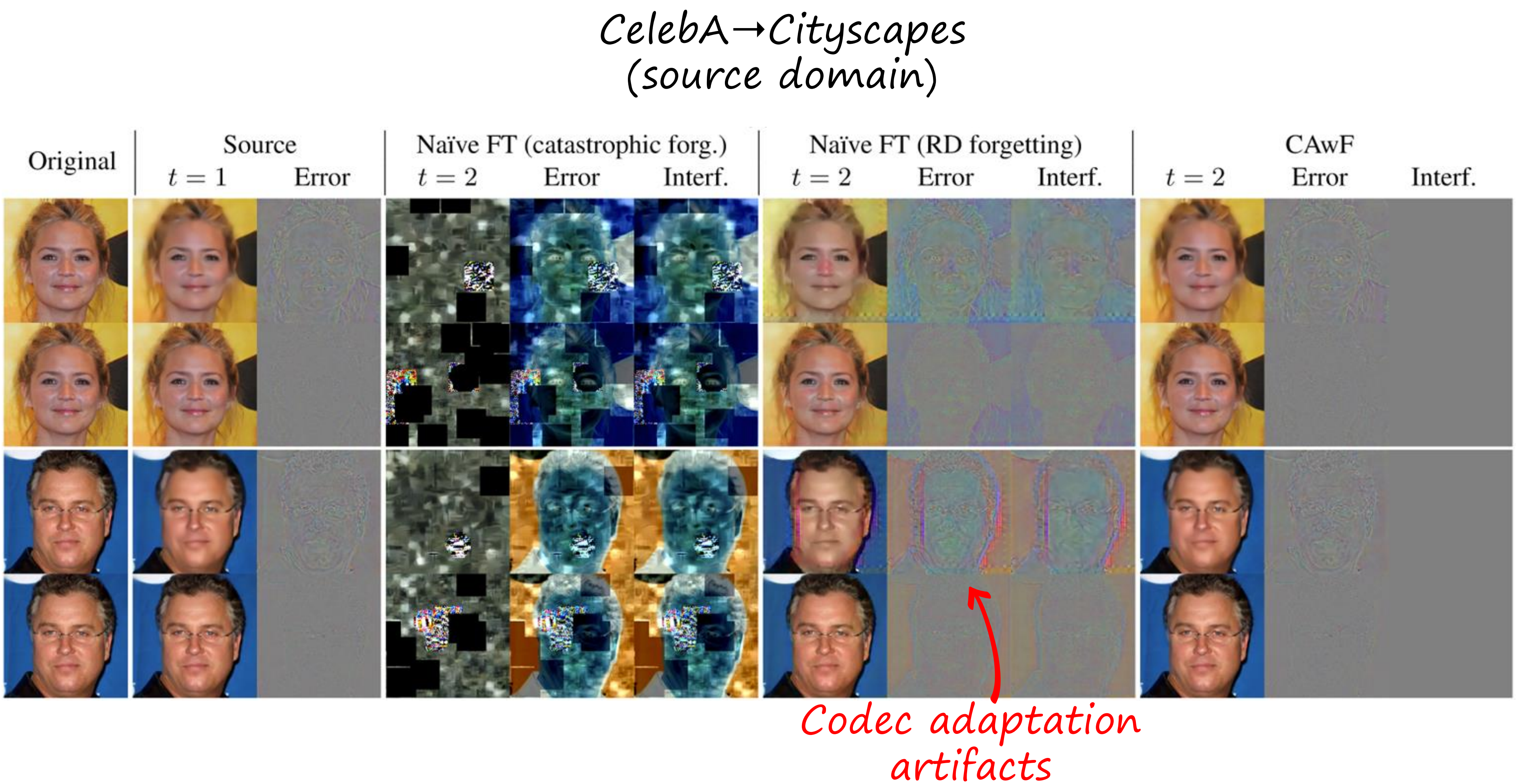
Codec adaptation without forgetting (CAwF)

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



DANICE: Domain adaptation without forgetting in neural image compression, CLIC 2021 at CVPR 2021

Codec adaptation without forgetting (CAwF)



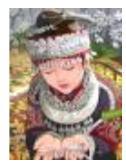
[DANICE: Domain adaptation without forgetting in neural image compression](), CLIC 2021 at CVPR 2021

Outline

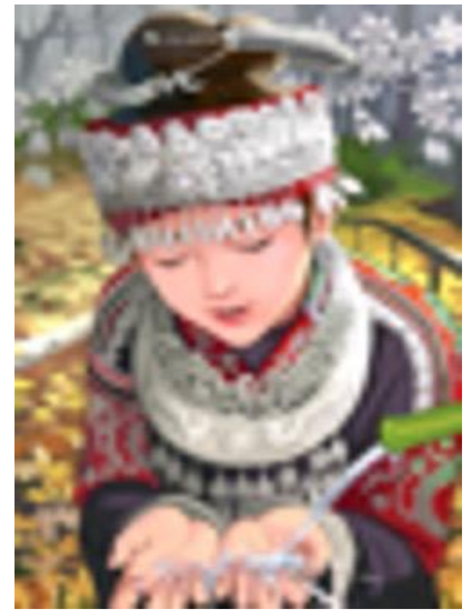
- Introduction: image/video coding
- Compression with neural networks
- Towards practical image compression
- Visual quality: perception vs distortion
- Video restoration and applications

Image superresolution

Downsampling
(25%)



Upsampling
(bicubic 4x)

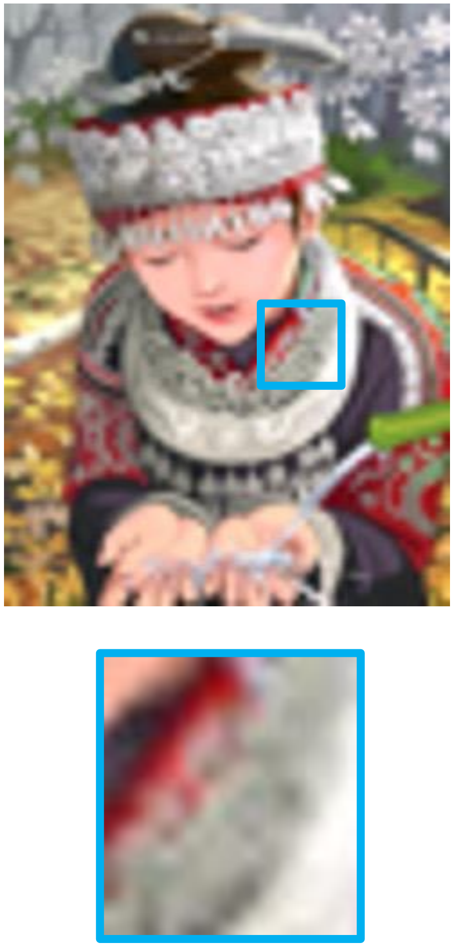


Note: lossy
(lost information
can't be recovered)

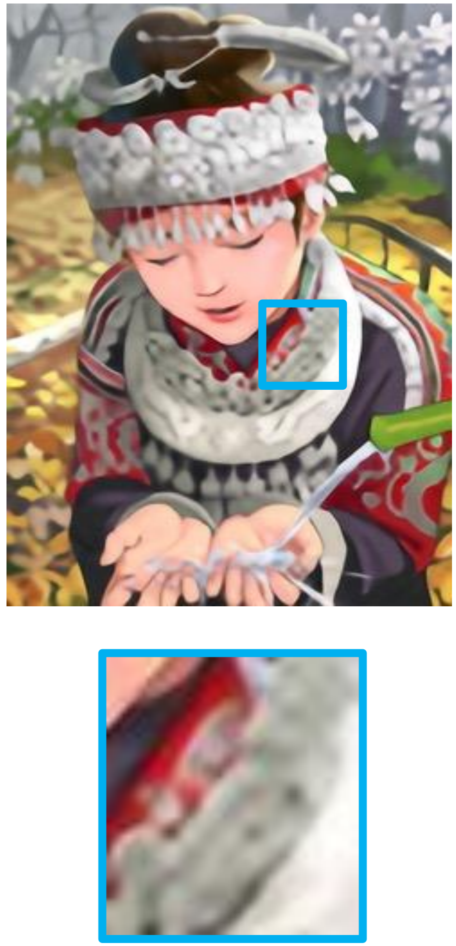
Image superresolution

Is (MSE/PSNR) distortion a good quality metric?

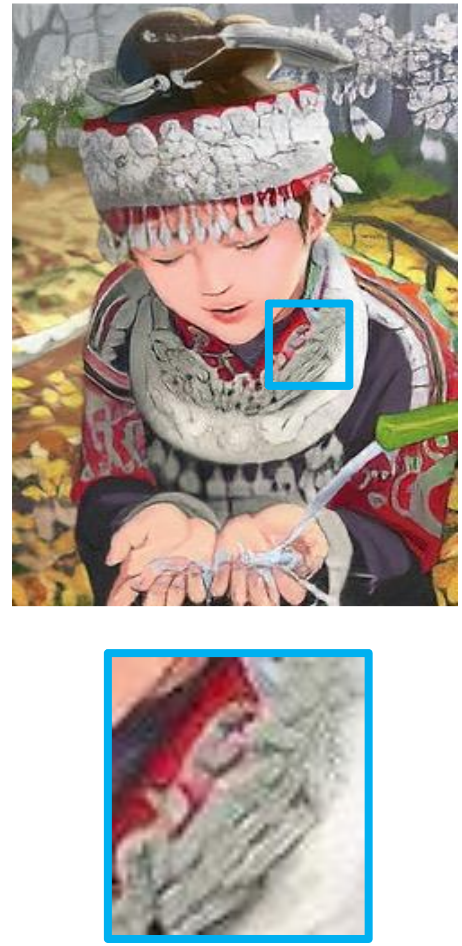
Bicubic
PNSR 21.59 dB



SRResNet (MSE)
PNSR 23.53 dB



SRGAN
PNSR 21.15 dB



Original

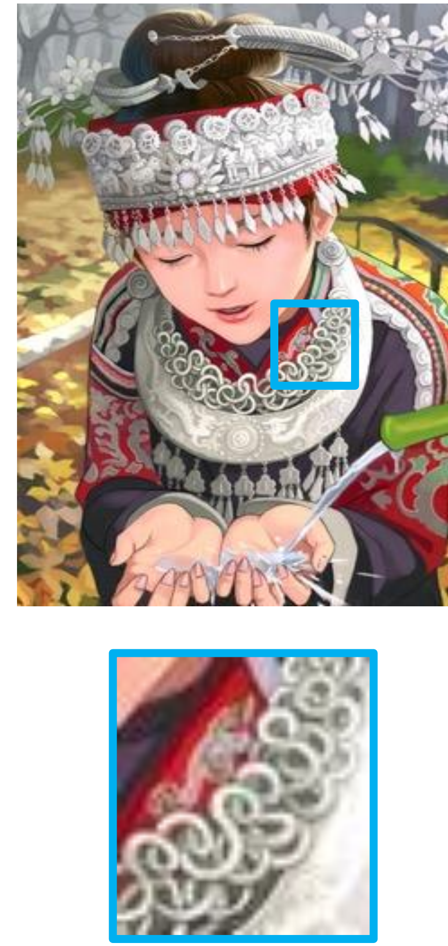
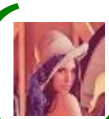


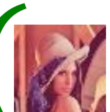
Image quality assessment: full-reference vs no-reference metrics

Distortion metric (full-reference)

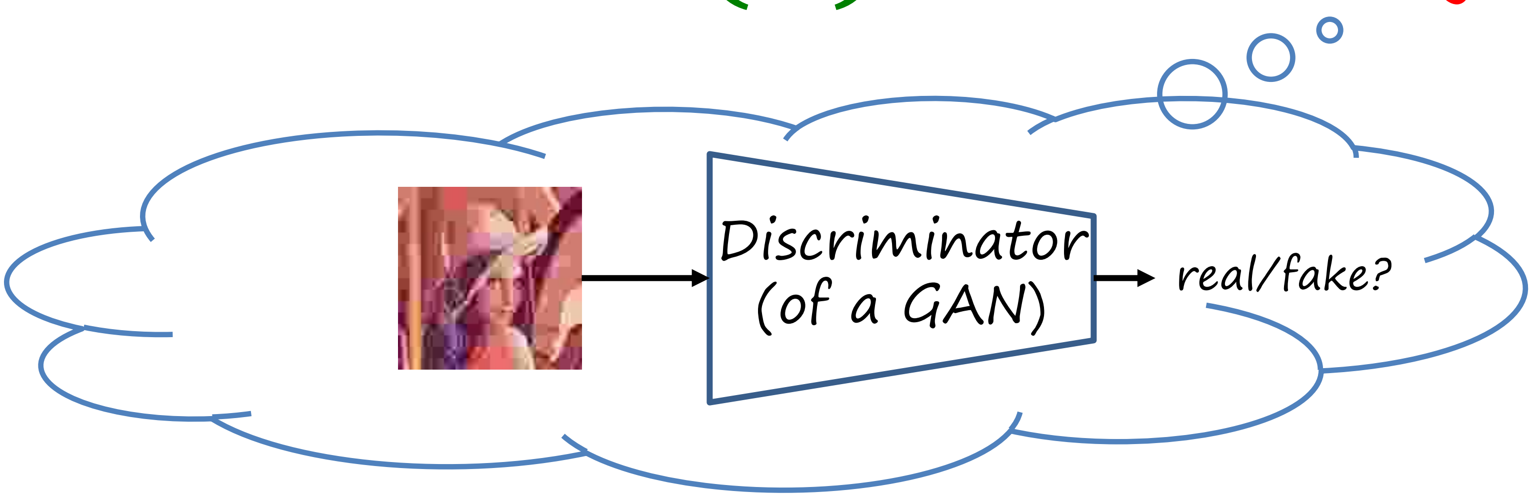
$D($  ,  $)$

How close is the image to the original one?

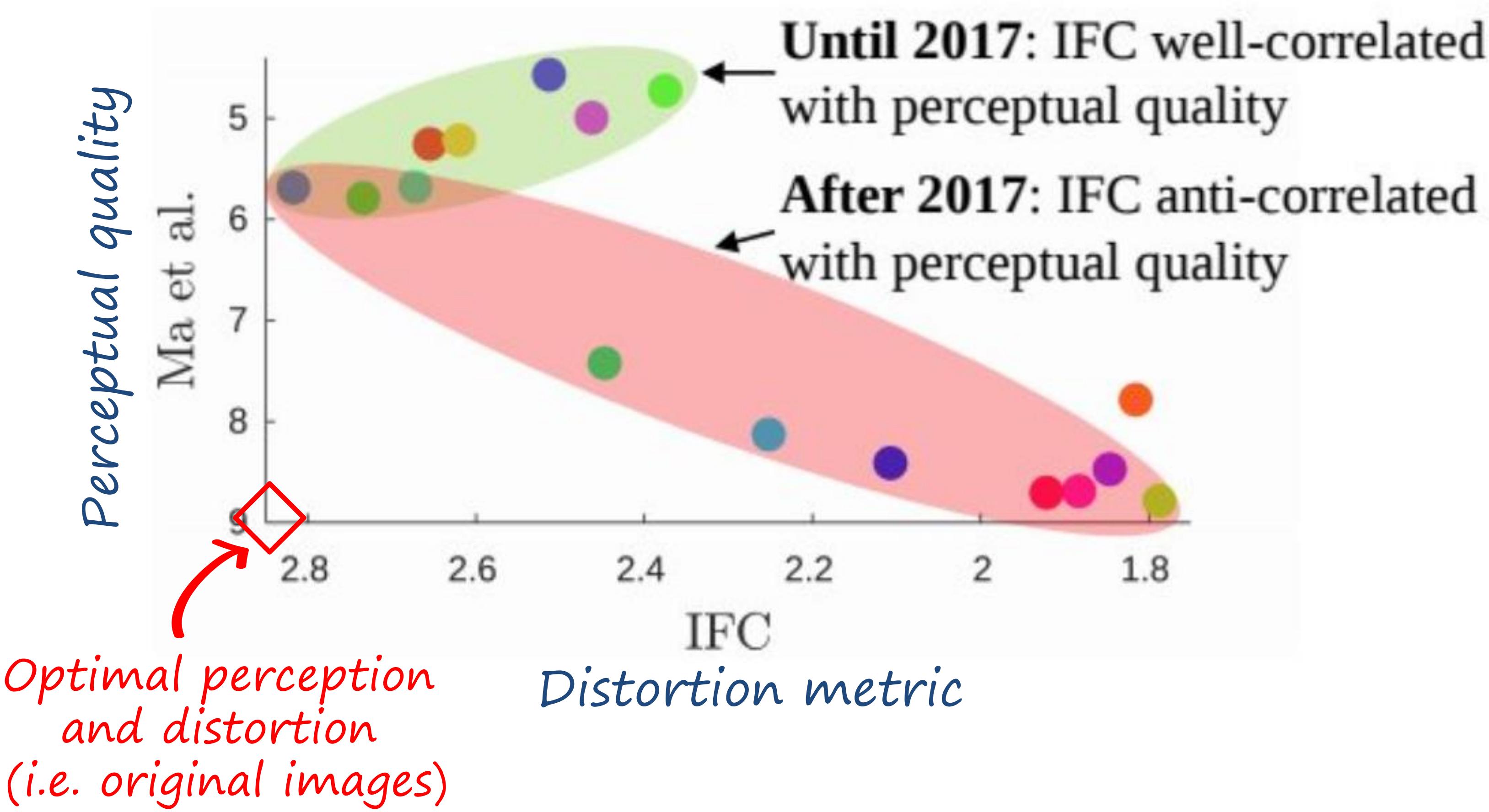
Perceptual metric (no-reference)

$P($  $)$

How realistic is the image?

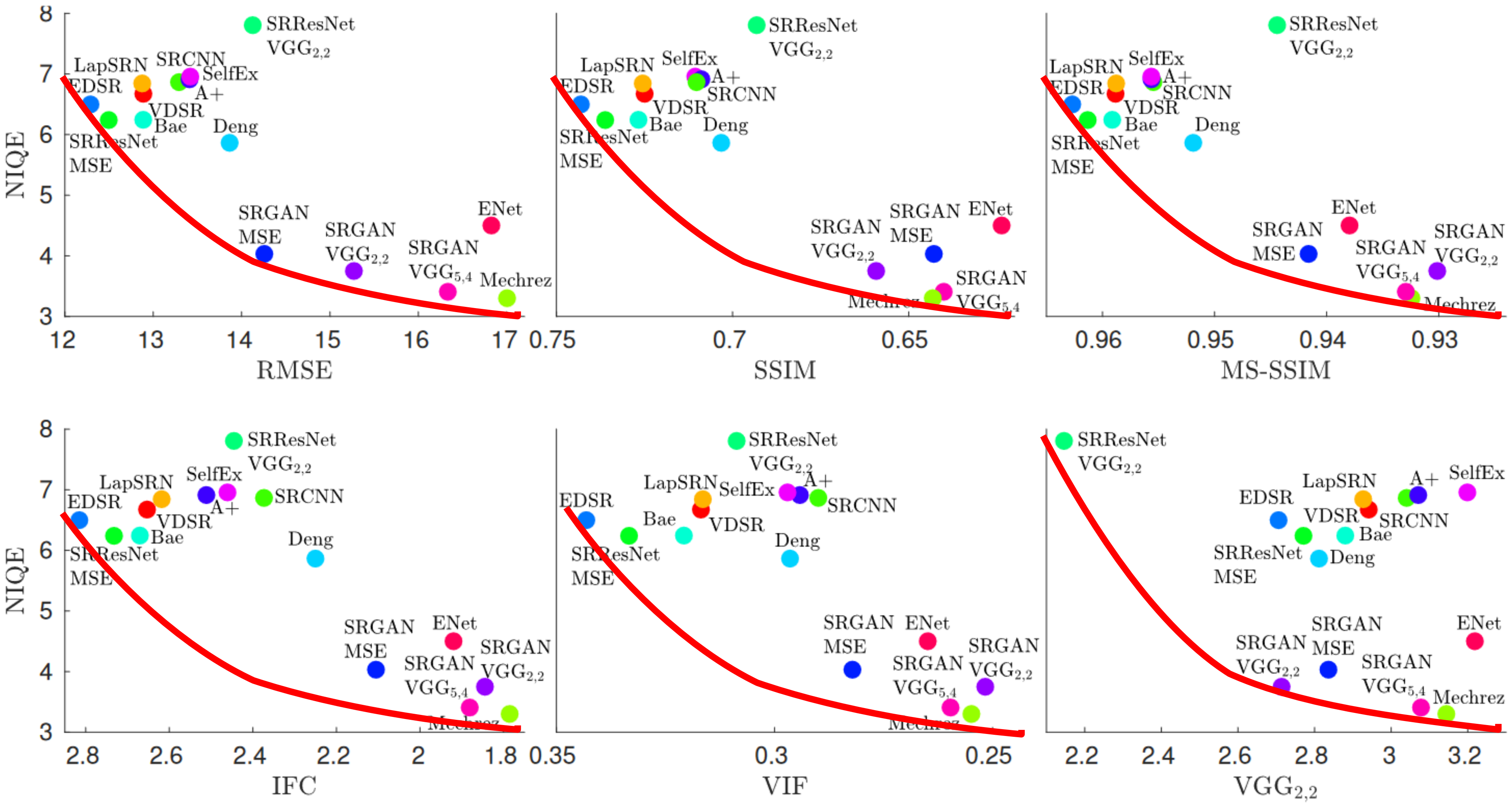


Perception-distortion in image superresolution methods



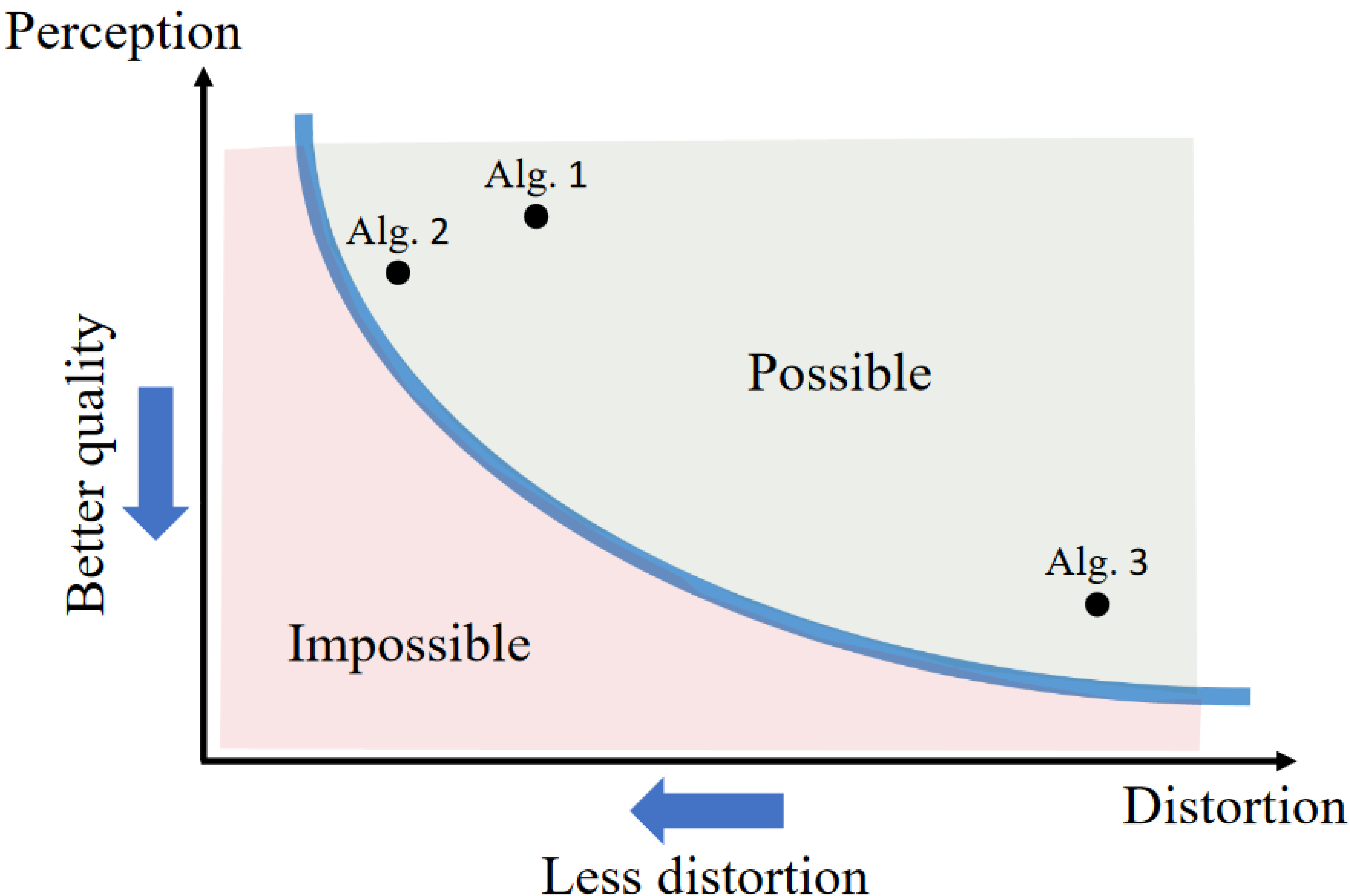
Slide adapted from Y. Blau

Perception-distortion in image superresolution methods



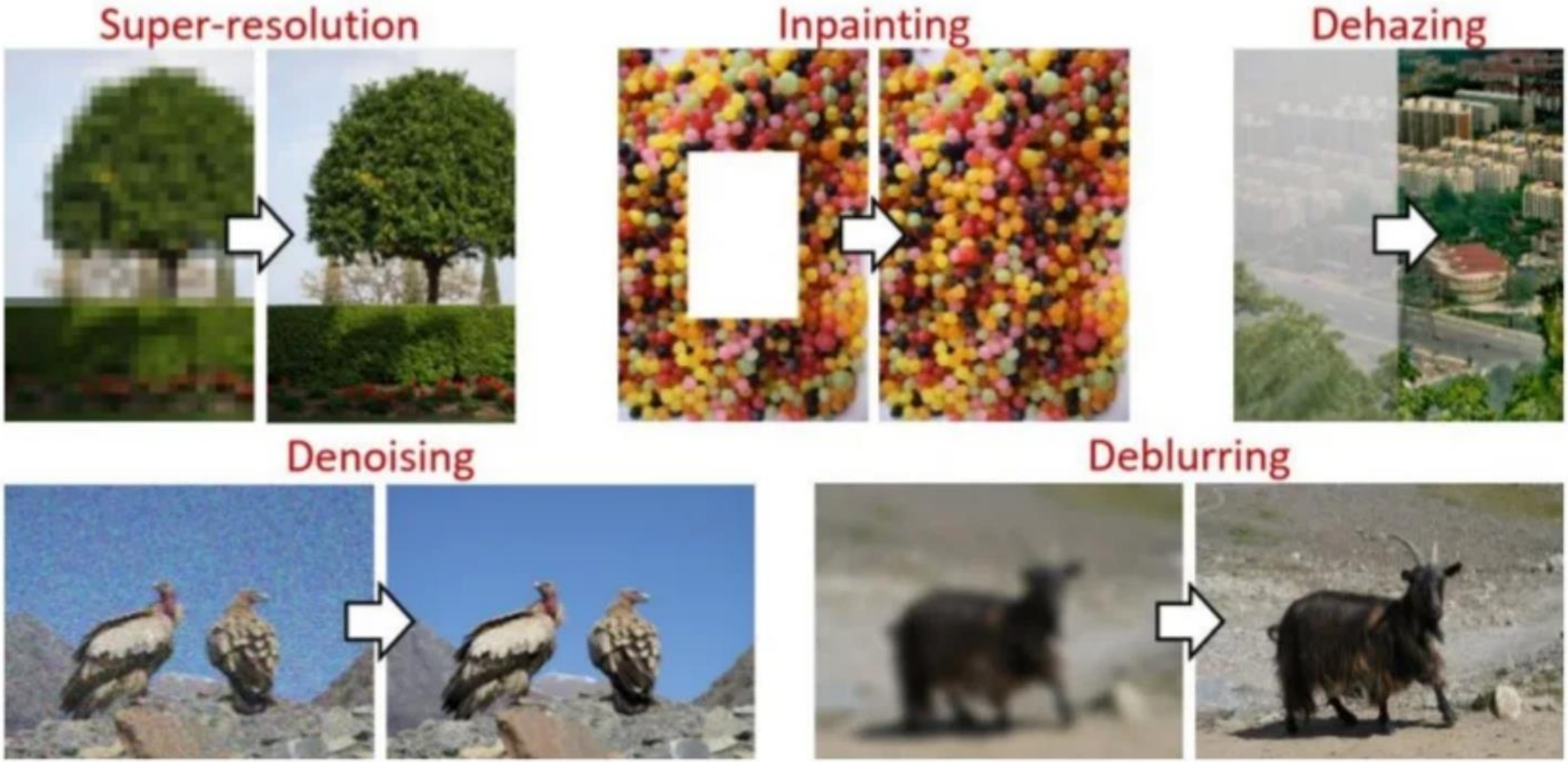
Slide credit: Y. Blau

Perception-distortion tradeoff



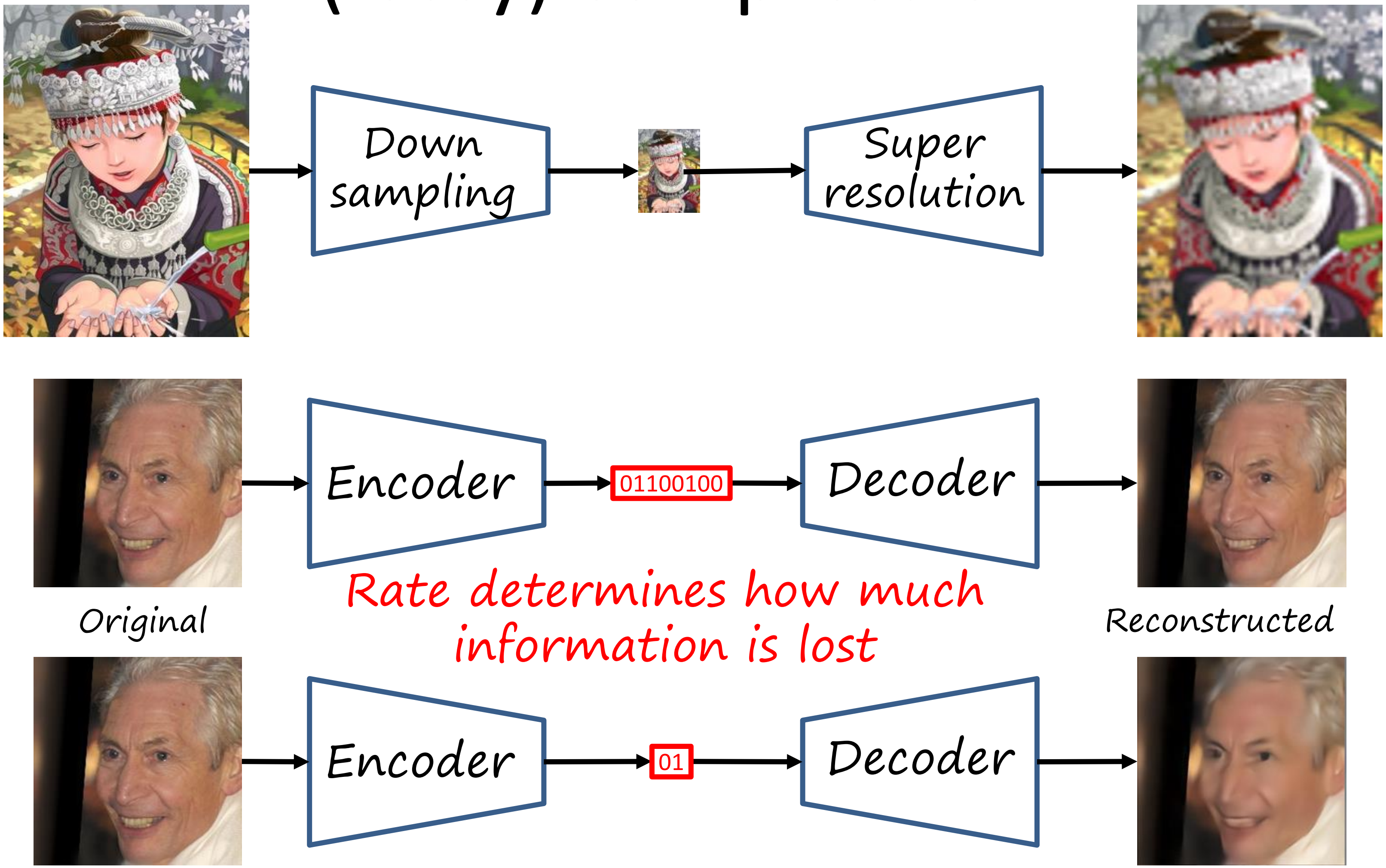
The Perception-Distortion Tradeoff, CVPR 2018

Image restoration problems

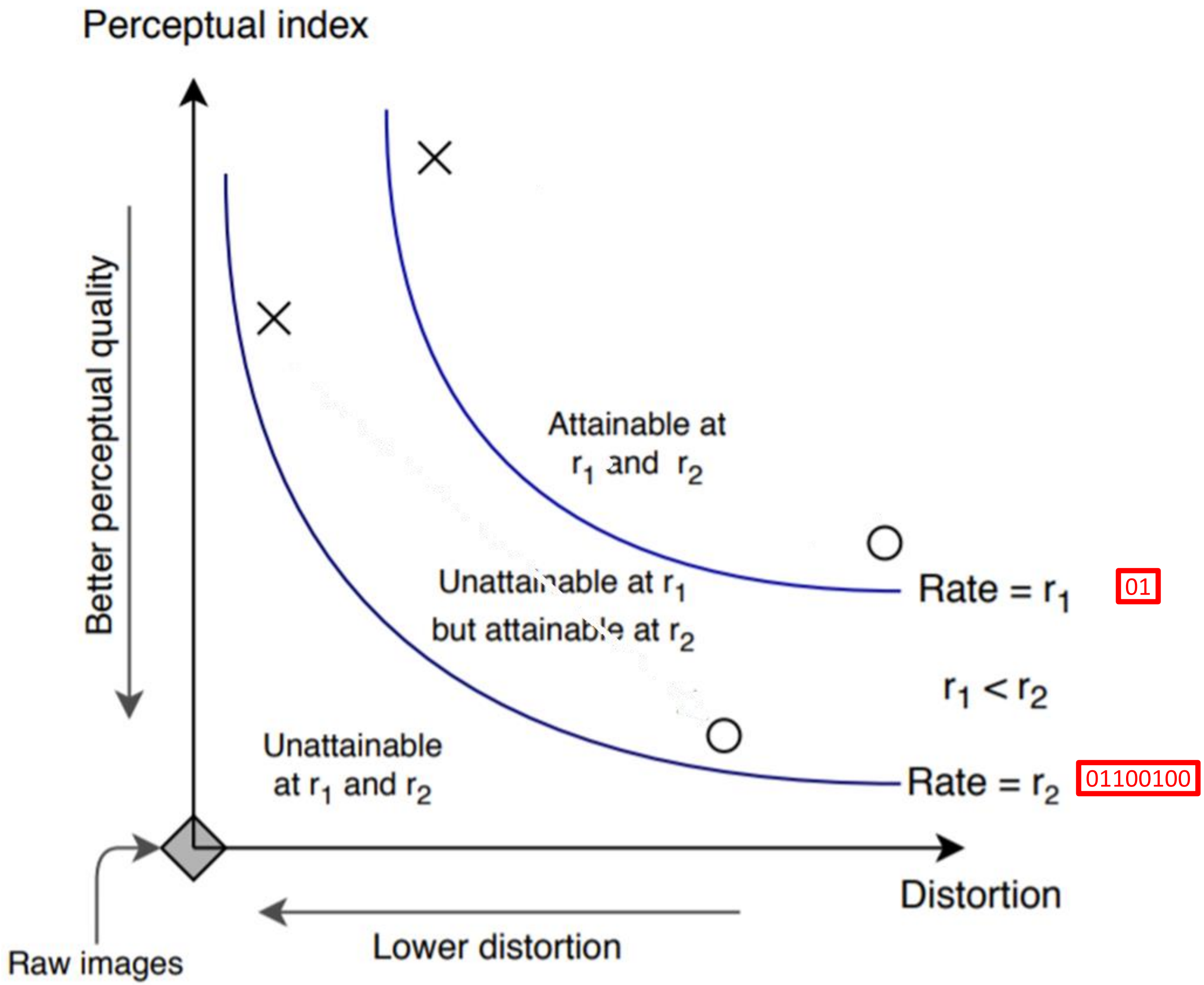


Slide credit: Y. Blau

What does this have to do with (lossy) compression?



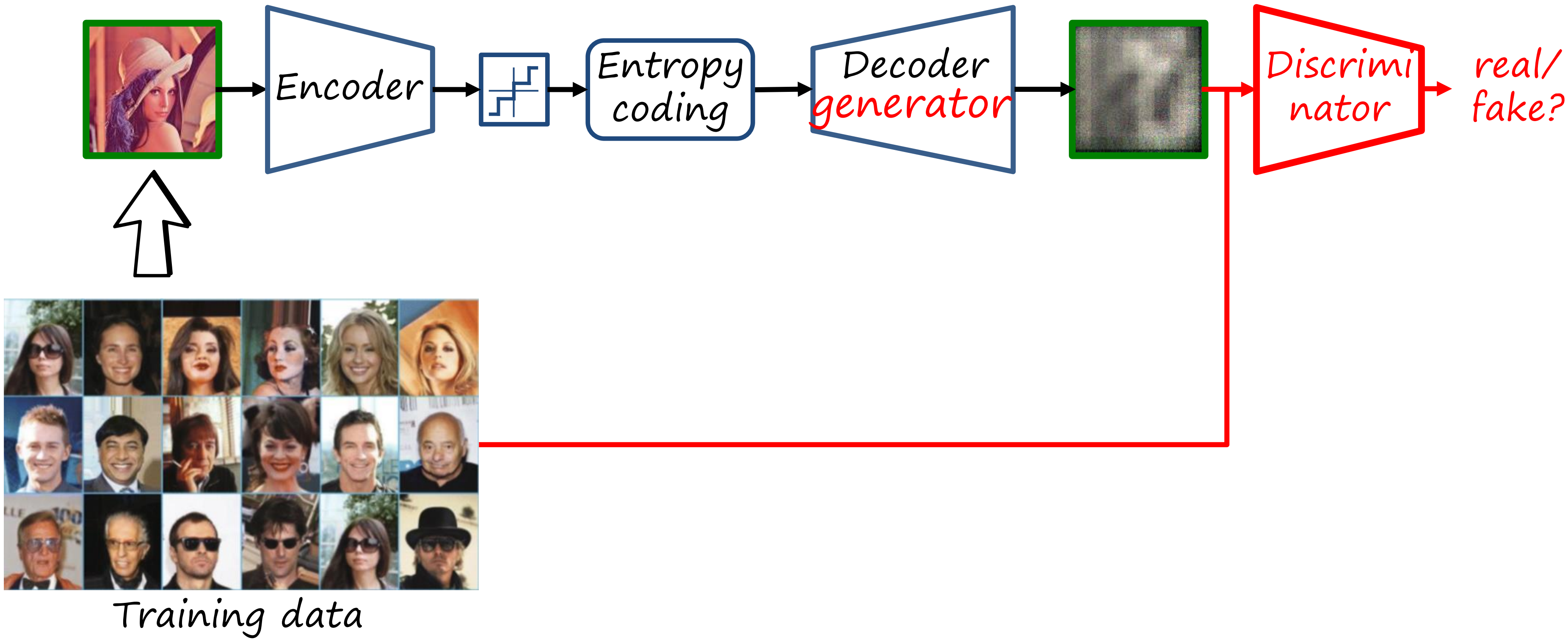
Rate-distortion-perception tradeoff



[Rethinking Lossy Compression: The Rate-Distortion-Perception Tradeoff](), ICML 2019

Generative lossy compression

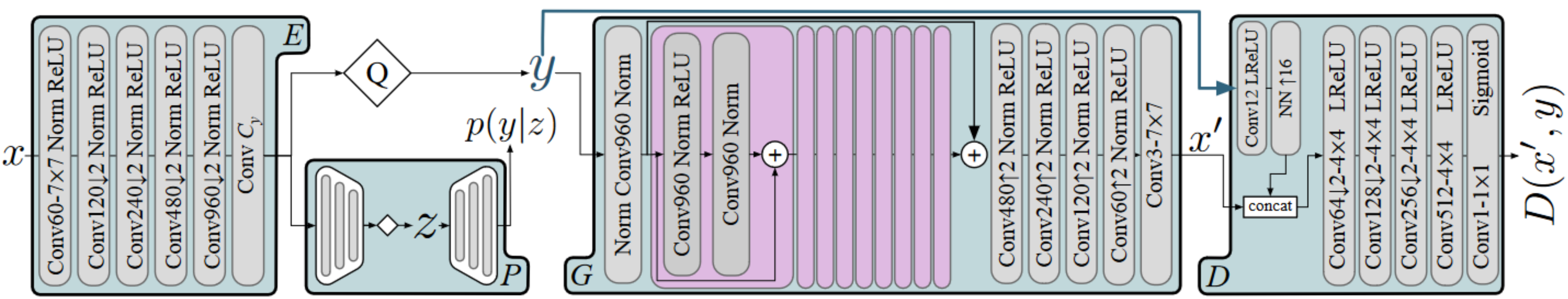
Optimize perception using a discriminator and adversarial loss
The decoder acts as generator of a conditional GAN



[High-Fidelity Generative Image Compression](), NeurIPS 2020

Generative lossy compression

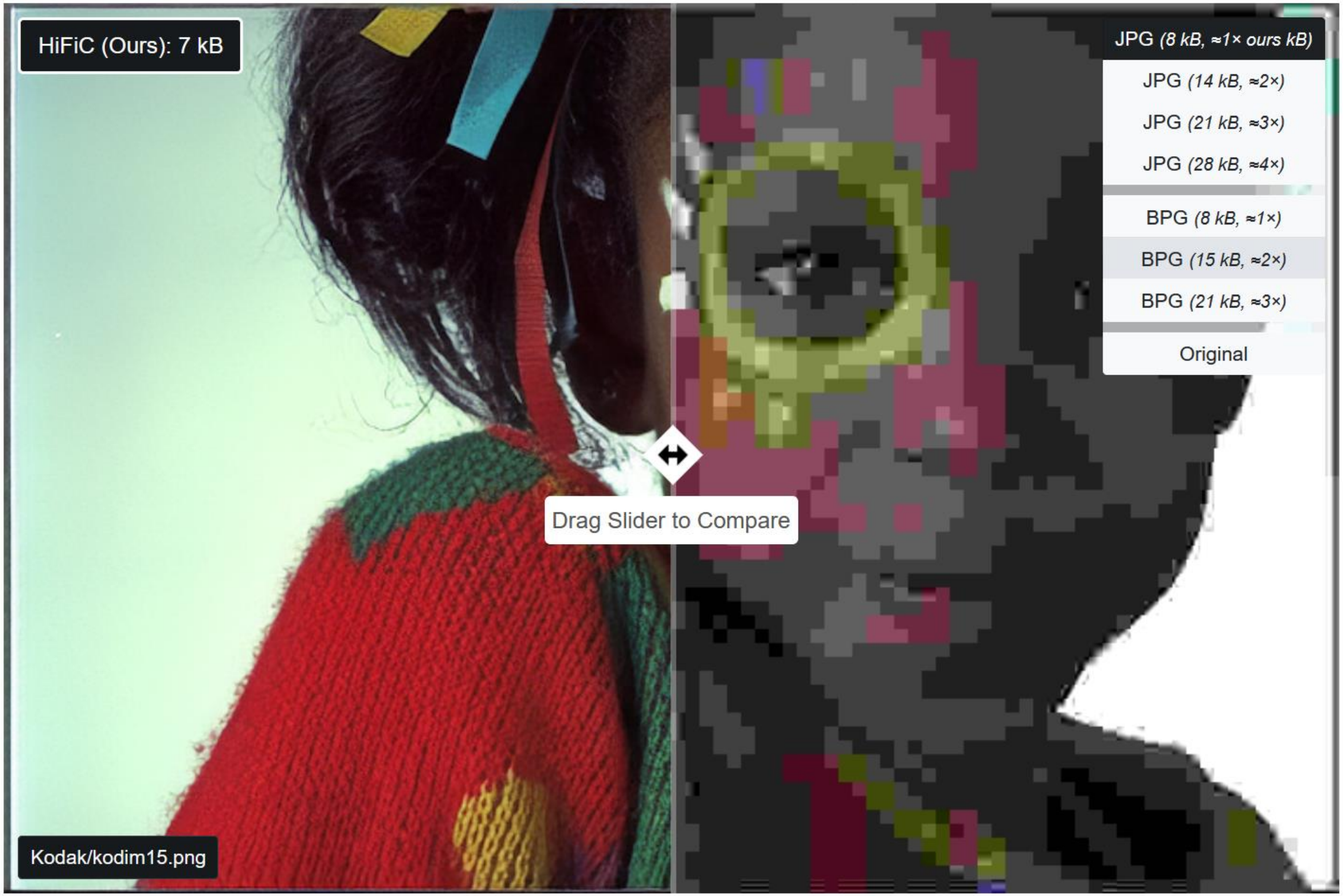
HiFiC: High-Fidelity Generative Image Compression



High-Fidelity Generative Image Compression, NeurIPS 2020

Generative lossy compression

HiFiC (7 kB) vs JPEG (8 kB)

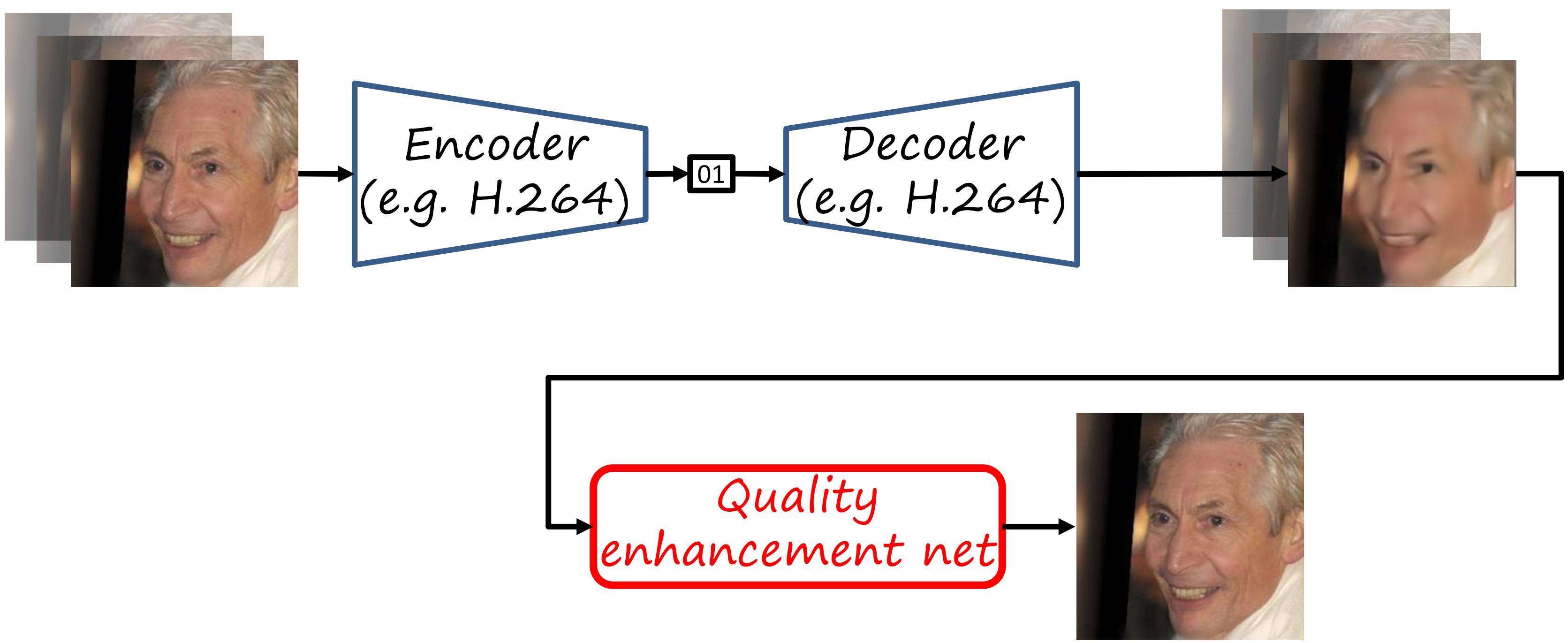


High-Fidelity Generative Image Compression, NeurIPS 2020

Outline

- Introduction: image/video coding
- Compression with neural networks
- Towards practical image compression
- Visual quality: perception vs distortion
- Video restoration and applications

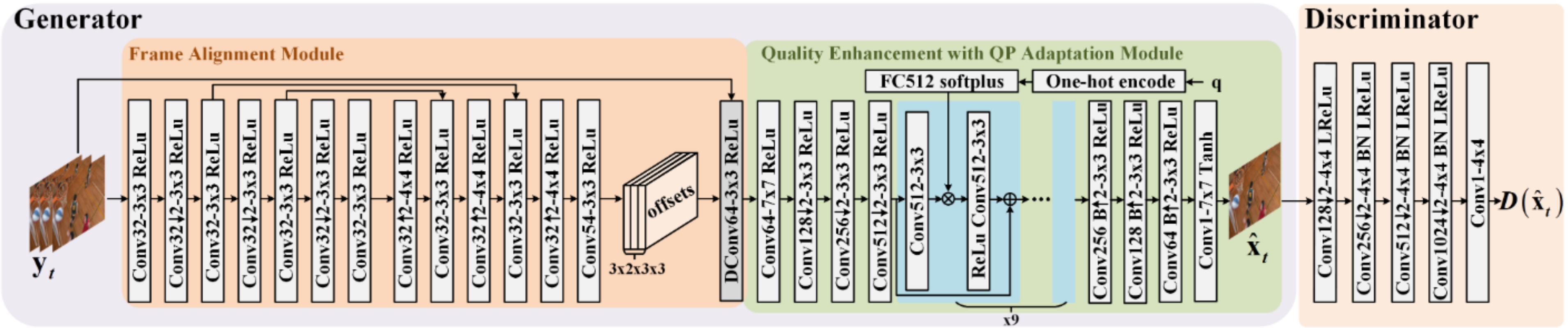
Video quality enhancement



Objectives:
- Align several frames
- Combine the aligned information

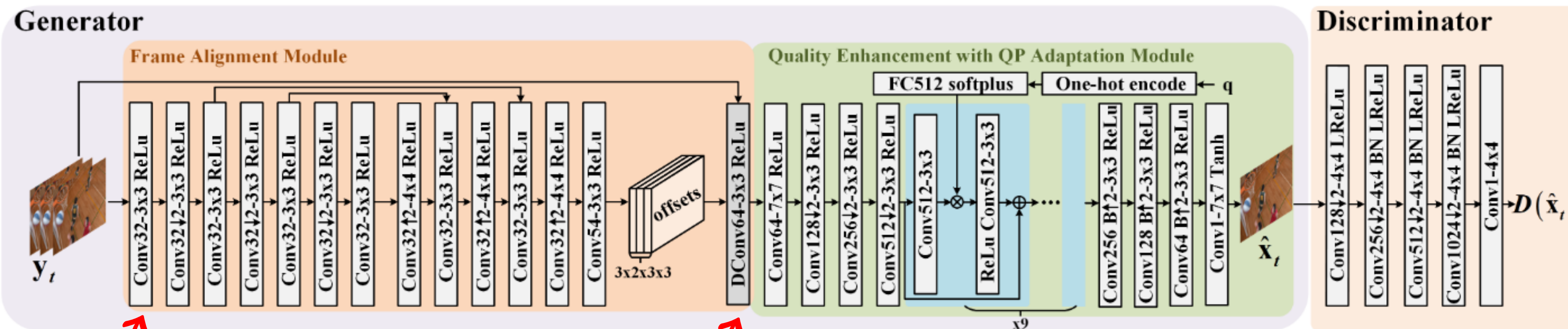
DCNGAN

Improvements:
- Use deformable convoution for alignment
- Condition on quantization parameter QP

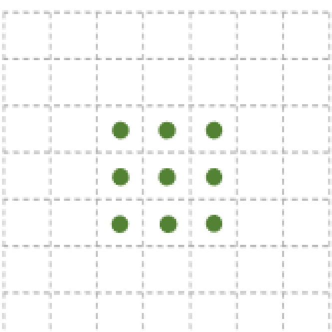


[DCNGAN: A deformable convolution-based GAN with QP adaptation for perceptual quality enhancement of compressed video](), ICASSP 2022

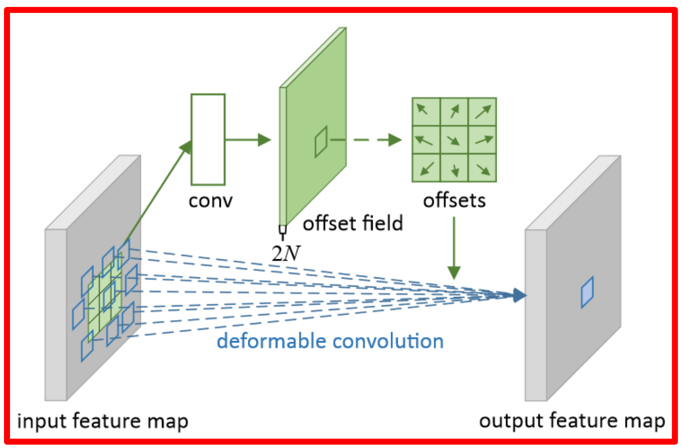
DCNGAN



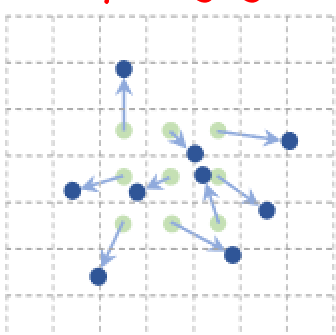
Convolution
Sampling grid



Deformable convolution

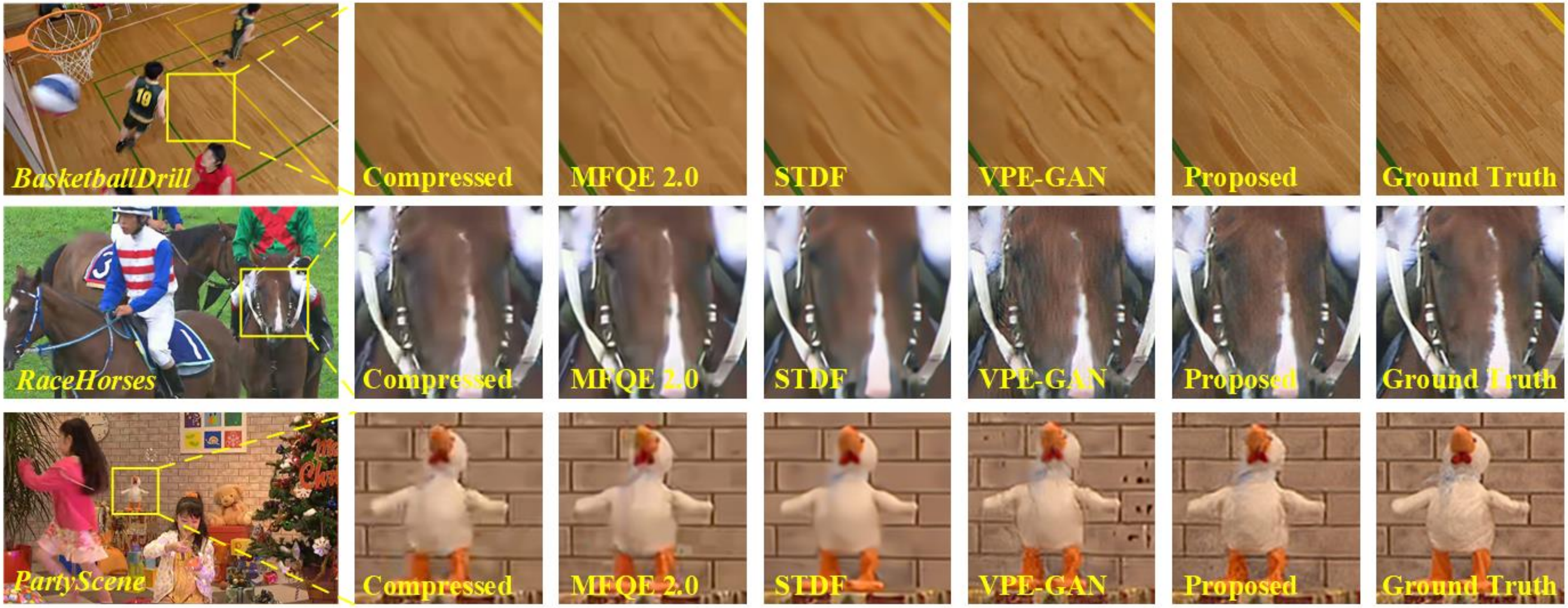


Sampling grid



Deformable Convolutional Networks, ICCV 2017
DCNGAN: A deformable convolution-based GAN with QP adaptation for perceptual quality enhancement of compressed video, ICASSP 2022

DCNGAN. Examples



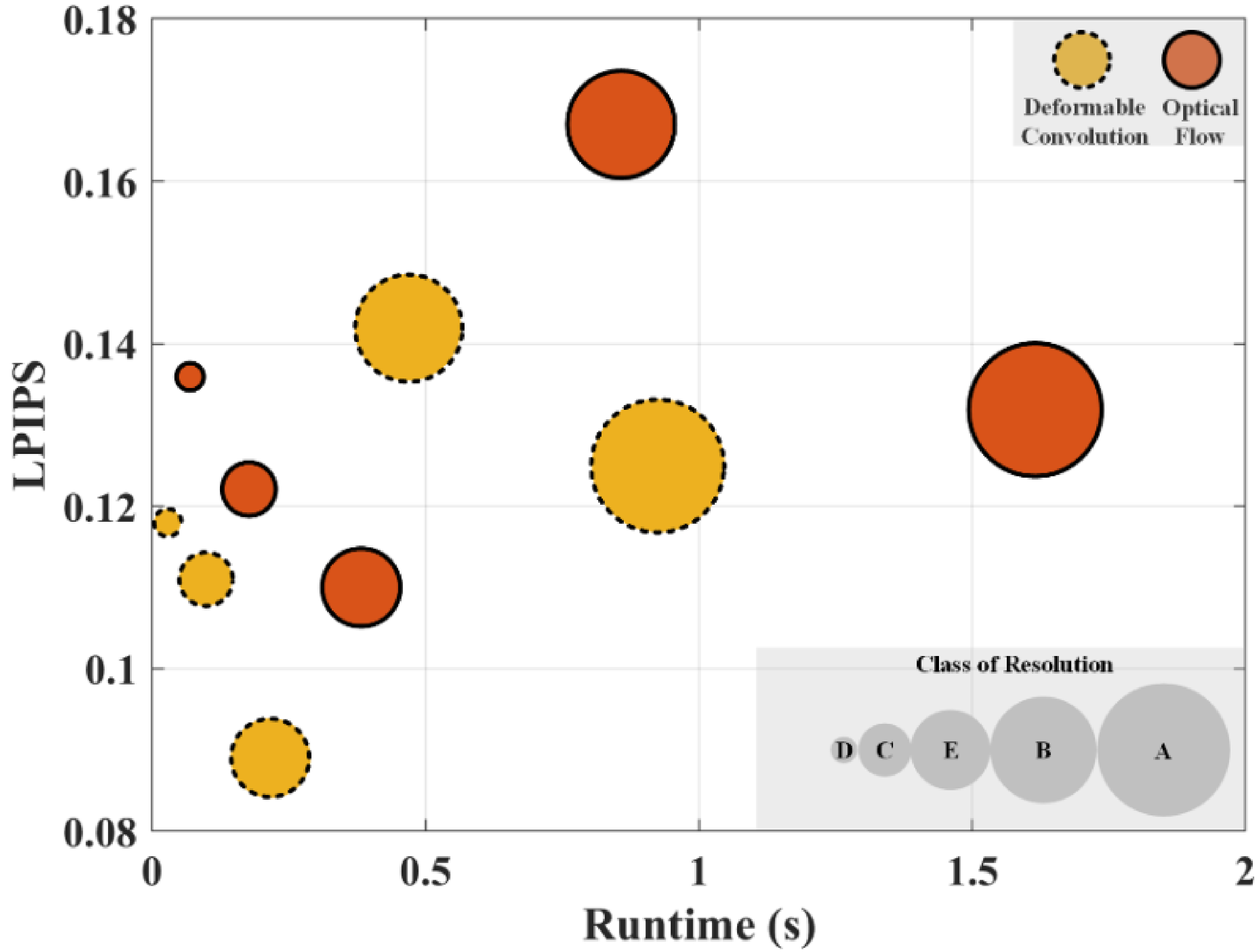
DCNGAN: A deformable convolution-based GAN with QP adaptation for perceptual quality enhancement of compressed video, ICASSP 2022

Video compression

QP		Sequences	Compressed LPIPS	Compressed DISTS	MFQE 2.0 [4] LPIPS	MFQE 2.0 [4] DISTS	STDF [5] LPIPS	STDF [5] DISTS	MW-GAN [9] LPIPS	MW-GAN [9] DISTS	VPE-GAN [10] LPIPS	VPE-GAN [10] DISTS	Proposed LPIPS	Proposed DISTS
32	*Class A*	*Traffic*	0.170	0.014	0.184	0.014	0.094	0.009	0.138	—	0.179	0.029	**0.070**	**0.006**
		PeopleOnStreet	0.150	0.018	0.167	0.018	0.133	0.010	0.130	—	0.135	0.015	**0.086**	**0.008**
	Class B	*Kimono*	0.258	0.043	0.294	0.046	0.160	0.026	0.189	—	0.180	0.034	**0.108**	**0.023**
		ParkScene	0.276	0.044	0.286	0.045	0.182	0.027	0.244	—	0.196	0.037	**0.123**	**0.023**
		Cactus	0.260	0.022	0.288	0.022	0.136	0.012	0.151	—	0.126	0.017	**0.096**	**0.010**
		BQTerrace	0.215	0.032	0.241	0.034	0.152	0.021	0.116	—	0.140	0.040	**0.113**	**0.018**
		BasketballDrive	0.247	0.028	0.279	0.031	0.166	0.022	0.141	—	0.132	0.025	**0.099**	**0.015**
	Class C	*RaceHorses*	0.147	0.066	0.174	0.075	0.120	0.061	0.126	—	0.101	0.055	**0.089**	**0.042**
		BQMall	0.124	0.066	0.145	0.071	0.089	0.050	0.091	—	0.112	0.063	**0.072**	**0.038**
		PartyScene	0.101	0.057	0.126	0.060	0.067	0.042	**0.026**	—	0.091	0.045	0.075	**0.029**
		BasketballDrill	0.156	0.073	0.181	0.079	0.126	0.068	0.109	—	0.105	0.060	**0.072**	**0.040**
	Class D	*RaceHorses*	0.122	0.121	0.143	0.132	0.098	0.113	0.117	—	0.093	0.126	**0.072**	**0.091**
		BQSquare	0.110	0.150	0.121	0.160	0.084	0.130	0.073	—	**0.066**	**0.112**	0.104	0.123
		BlowingBubbles	0.102	0.117	0.111	0.128	0.068	0.104	**0.063**	—	0.072	0.096	0.065	**0.084**
		BasketballPass	0.116	0.135	0.135	0.150	0.099	0.127	0.095	—	0.085	0.116	**0.067**	**0.099**
	Class E	*FourPeople*	0.120	0.037	0.128	0.038	0.089	0.022	0.080	—	0.103	0.028	**0.054**	**0.016**
		Johnny	0.148	0.035	0.159	0.035	0.111	0.021	0.083	—	0.178	0.059	**0.063**	**0.014**
		KristenAndSara	0.134	0.038	0.148	0.039	0.106	0.025	0.108	—	0.136	0.046	**0.062**	**0.019**
		Average	0.164	0.061	0.184	0.065	0.116	0.049	0.115	—	0.124	0.056	**0.083**	**0.039**
22		*Average*	0.077	0.020	0.087	0.022	0.050	**0.014**	—	—	0.097	0.047	**0.042**	0.017
27		*Average*	0.116	0.037	0.130	0.040	0.077	0.029	—	—	0.103	0.054	**0.059**	**0.026**
37		*Average*	0.223	0.089	0.232	0.086	0.168	0.080	0.177	—	0.148	0.070	**0.120**	**0.058**

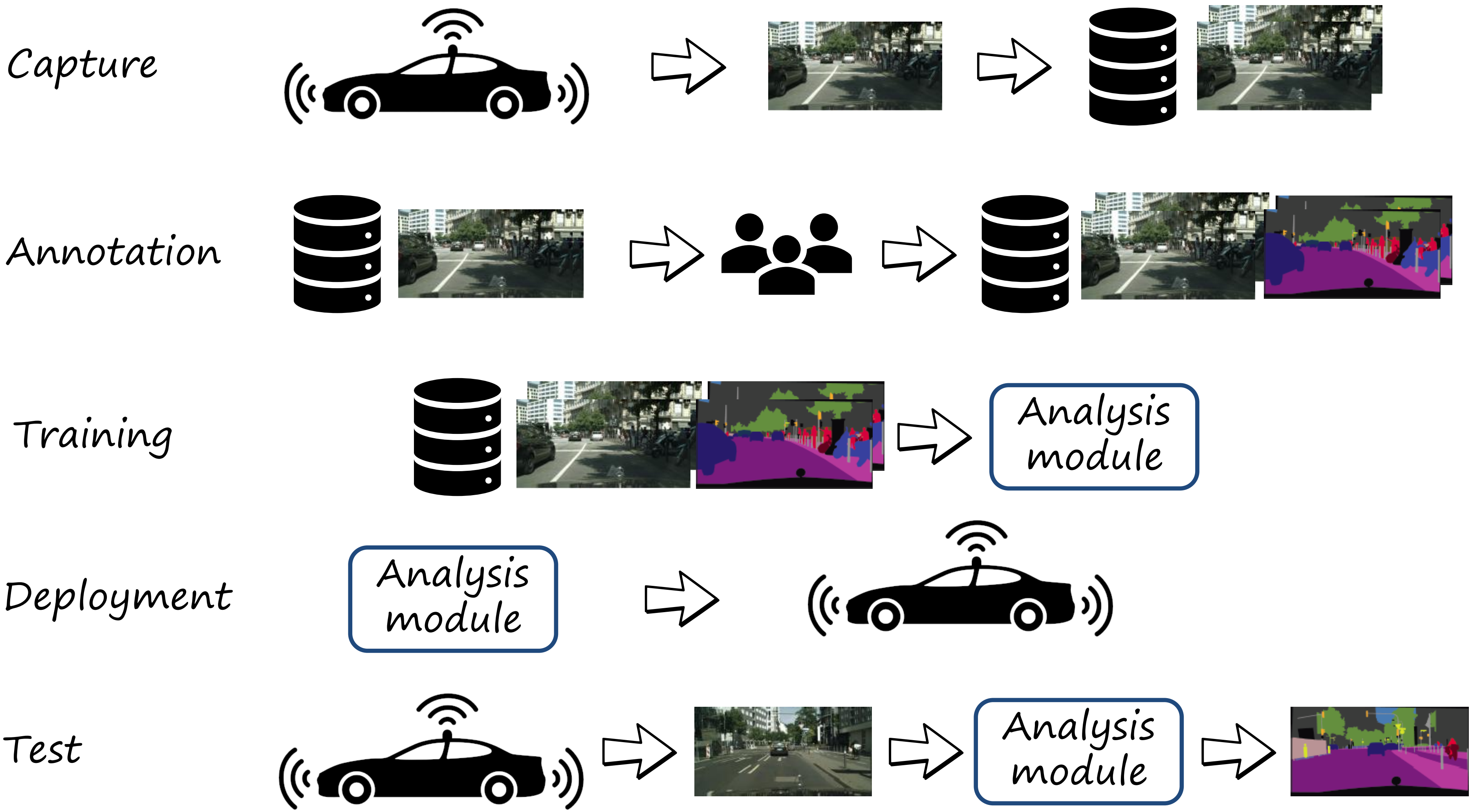
DCNGAN: A deformable convolution-based GAN with QP adaptation for perceptual quality enhancement of compressed video, ICASSP 2022

DConv vs optical flow



[DCNGAN: A deformable convolution-based GAN with QP adaptation for perceptual quality enhancement of compressed video](), ICASSP 2022

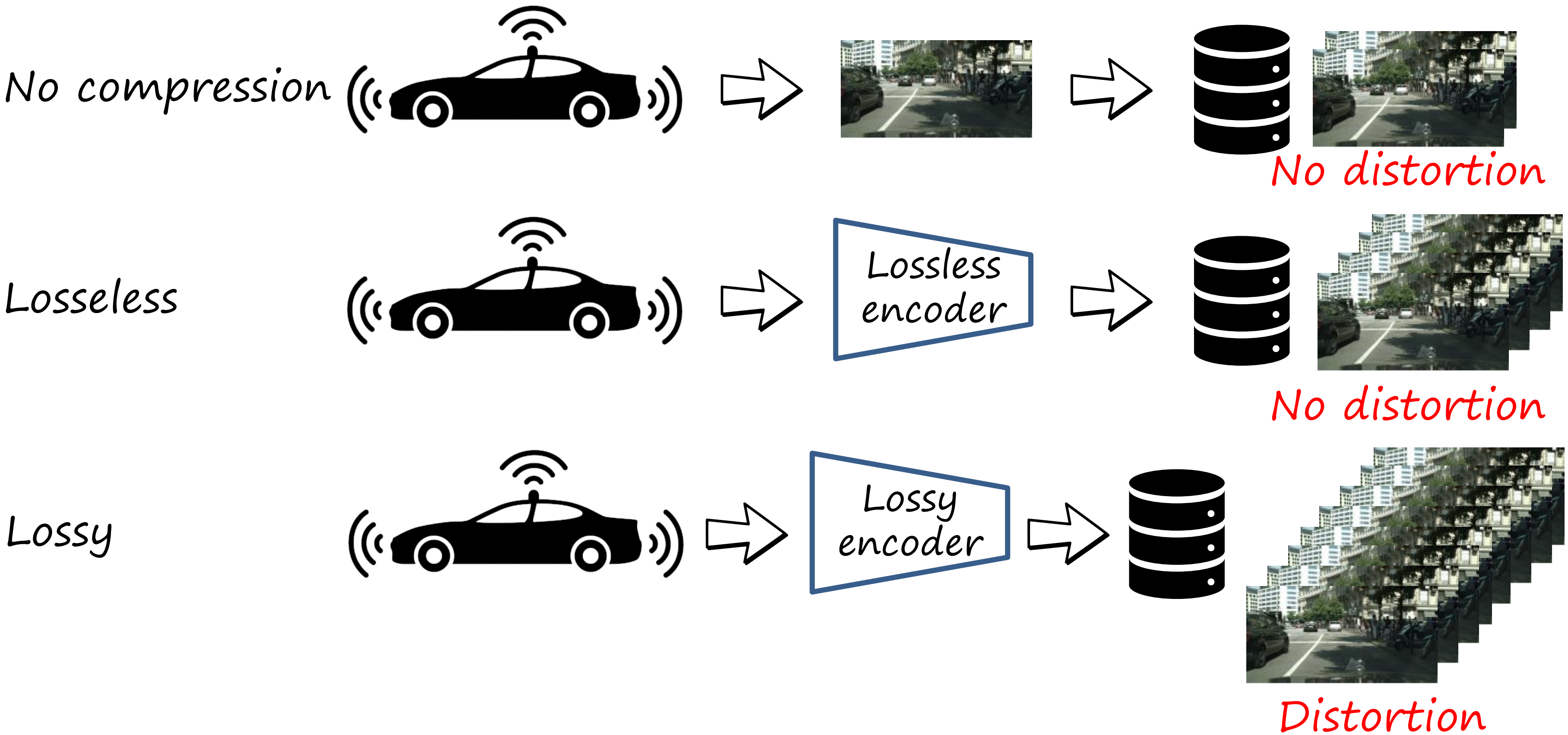
Data collection for onboard perception



The more images, the better model (in principle)

Distributed Learning and Inference with Compressed Images, IEEE Trans. Image Processing 2021

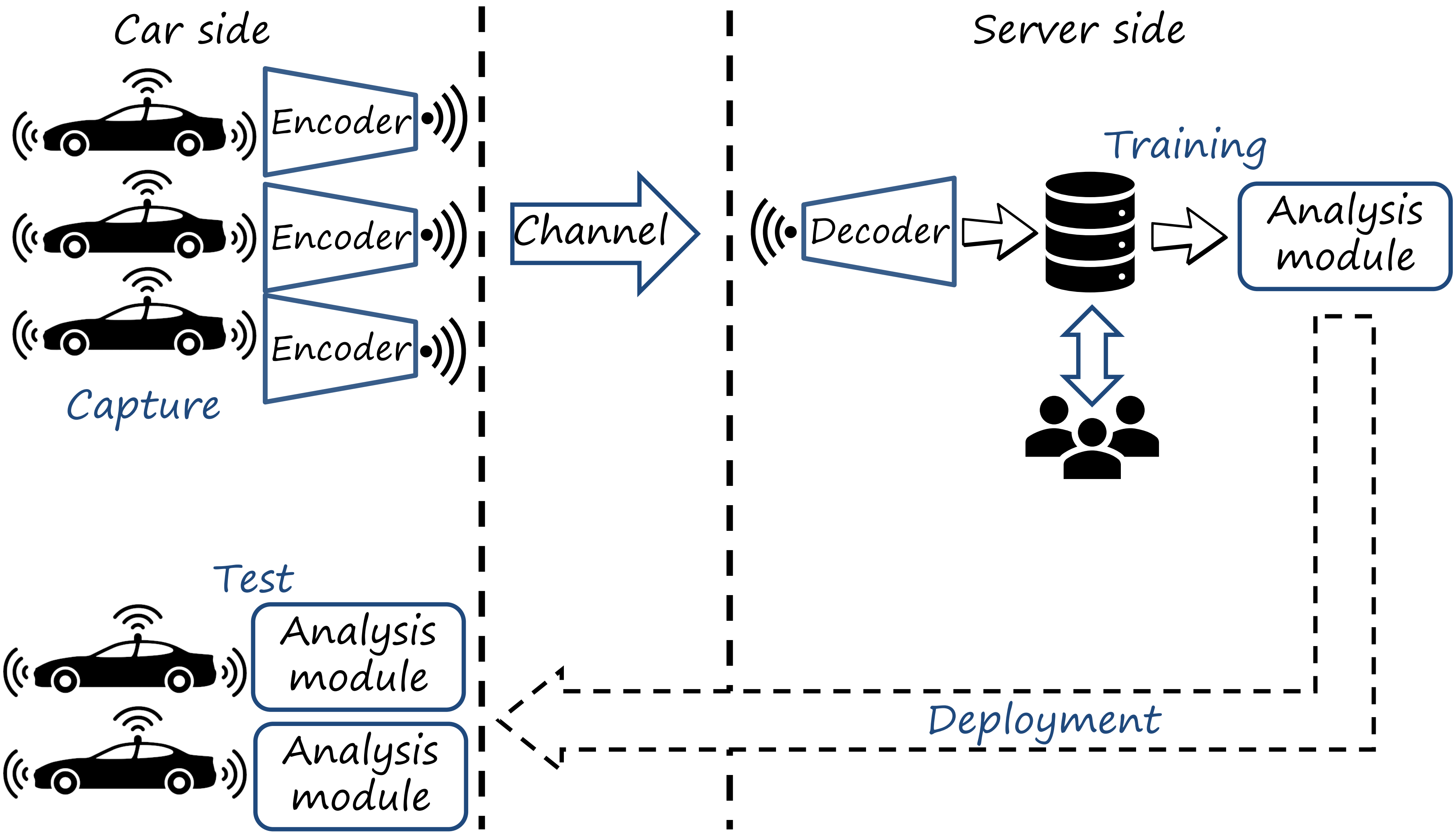
Data collection for onboard perception



The higher the compression rate the more images we can collect

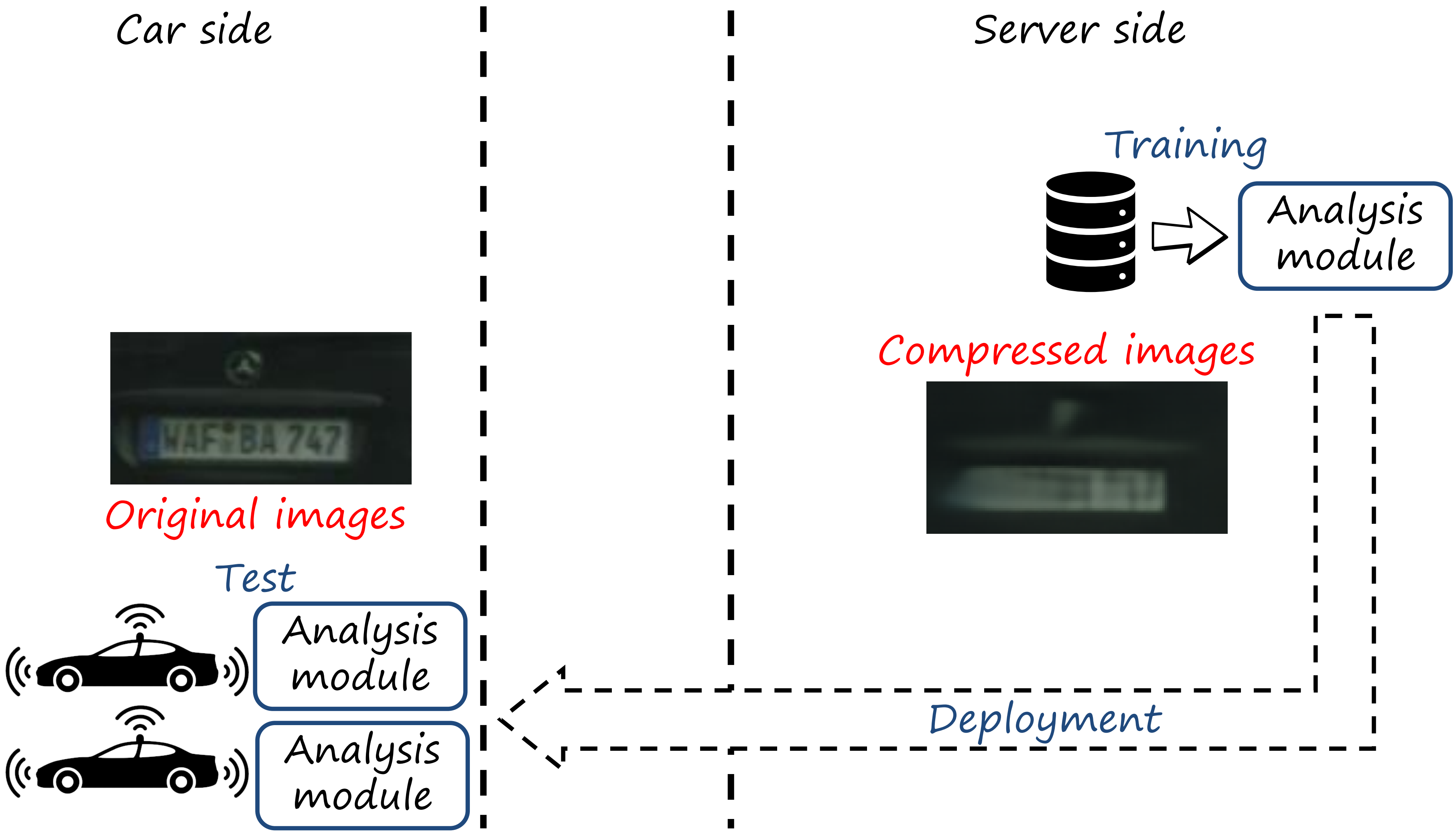
Distributed Learning and Inference with Compressed Images, IEEE Trans. Image Processing 2021

Distributed data collection



Distributed Learning and Inference with Compressed Images, IEEE Trans. Image Processing 2021

Distributed data collection



Distributed Learning and Inference with Compressed Images, IEEE Trans. Image Processing 2021

Training images vs test images

Training (compressed)

Test (original)

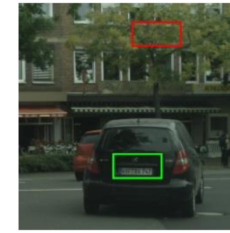
codec: mean-scale hyperprior

Distributed Learning and Inference with Compressed Images, IEEE Trans. Image Processing 2021

Training images vs test images

Training (compressed)

Test (original)



[Distributed Learning and Inference with Compressed Images](), IEEE Trans. Image Processing 2021

Training images vs test images

Training (compressed) Test (original)

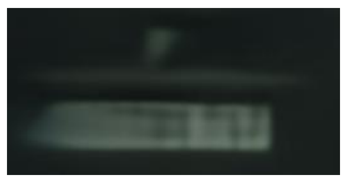
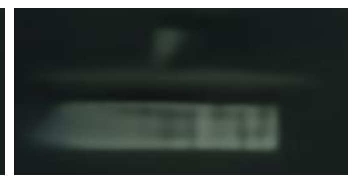
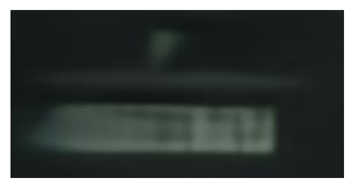
Configuration CO: compressed/original

Observation 1: training and test distributions are different (**covariate shift**)

Observation 2: training images have less information than test images (**loss of information**)

[Distributed Learning and Inference with Compressed Images](), IEEE Trans. Image Processing 2021

Training/test configurations

	Training	Test	Covariate shift	Information loss (training/test)
OO (ideal) original/original			No	No/No
CO compressed/ original			Yes	Yes/No
CC compressed/compressed			No	Yes/Yes
OC original/compressed			Yes	No/Yes

Distributed Learning and Inference with Compressed Images, IEEE Trans. Image Processing 2021

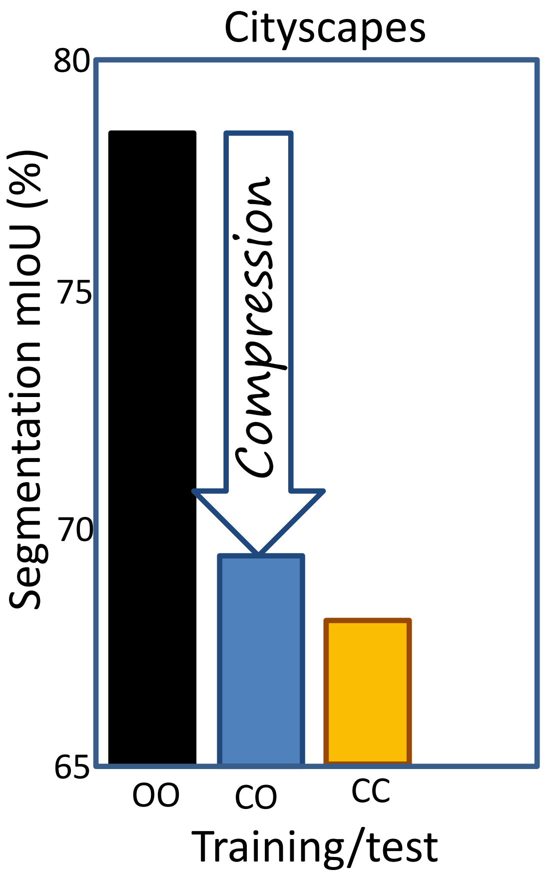
Effect on downstream task

Training Test

OO (ideal)

CO

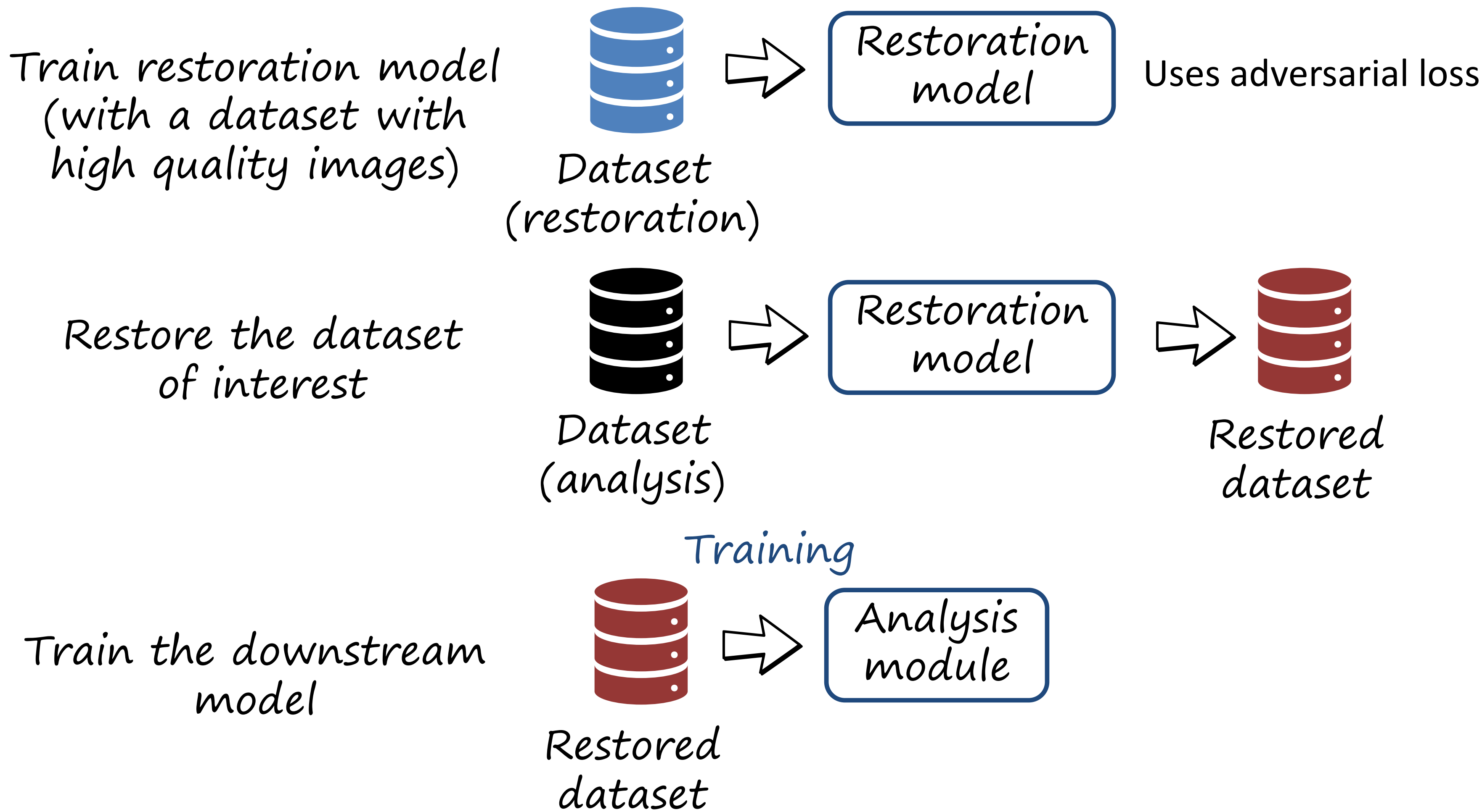
CC



Conclusion (this dataset): better to keep more information in test than reduce the covariate shift

[Distributed Learning and Inference with Compressed Images](), IEEE Trans. Image Processing 2021

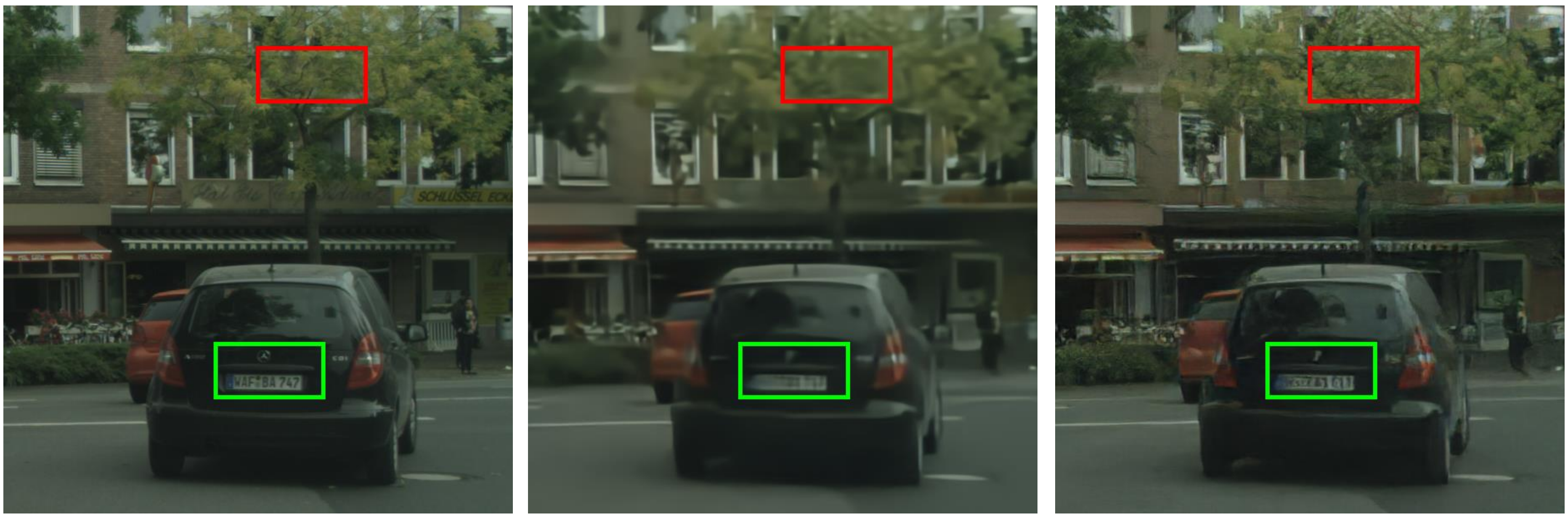
Proposed approach: dataset restoration



Distributed Learning and Inference with Compressed Images, IEEE Trans. Image Processing 2021

Training images vs test images

Original *Compressed* *Restored*



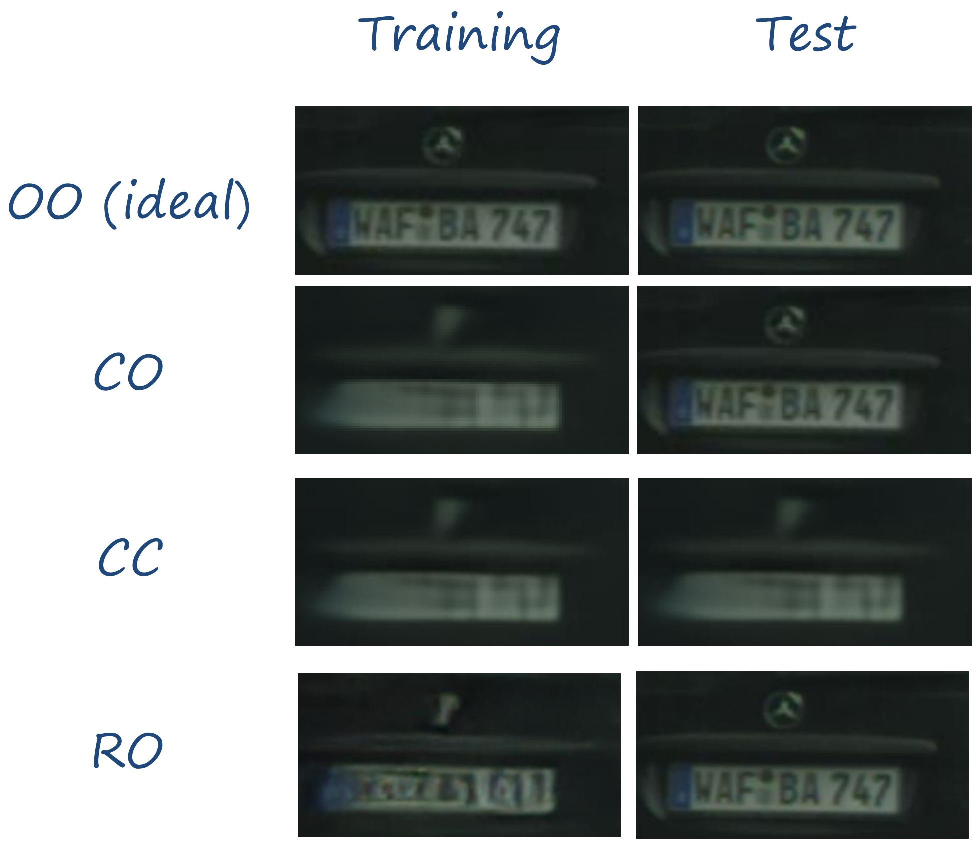
[Distributed Learning and Inference with Compressed Images](), IEEE Trans. Image Processing 2021

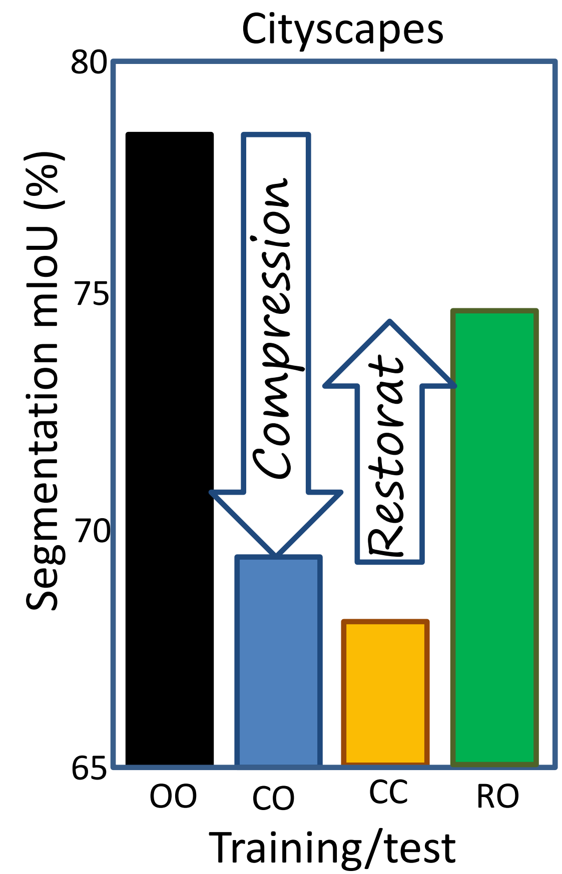
Training images vs test images

Original *Compressed* *Restored*

[Distributed Learning and Inference with Compressed Images](), IEEE Trans. Image Processing 2021

Effect on downstream task



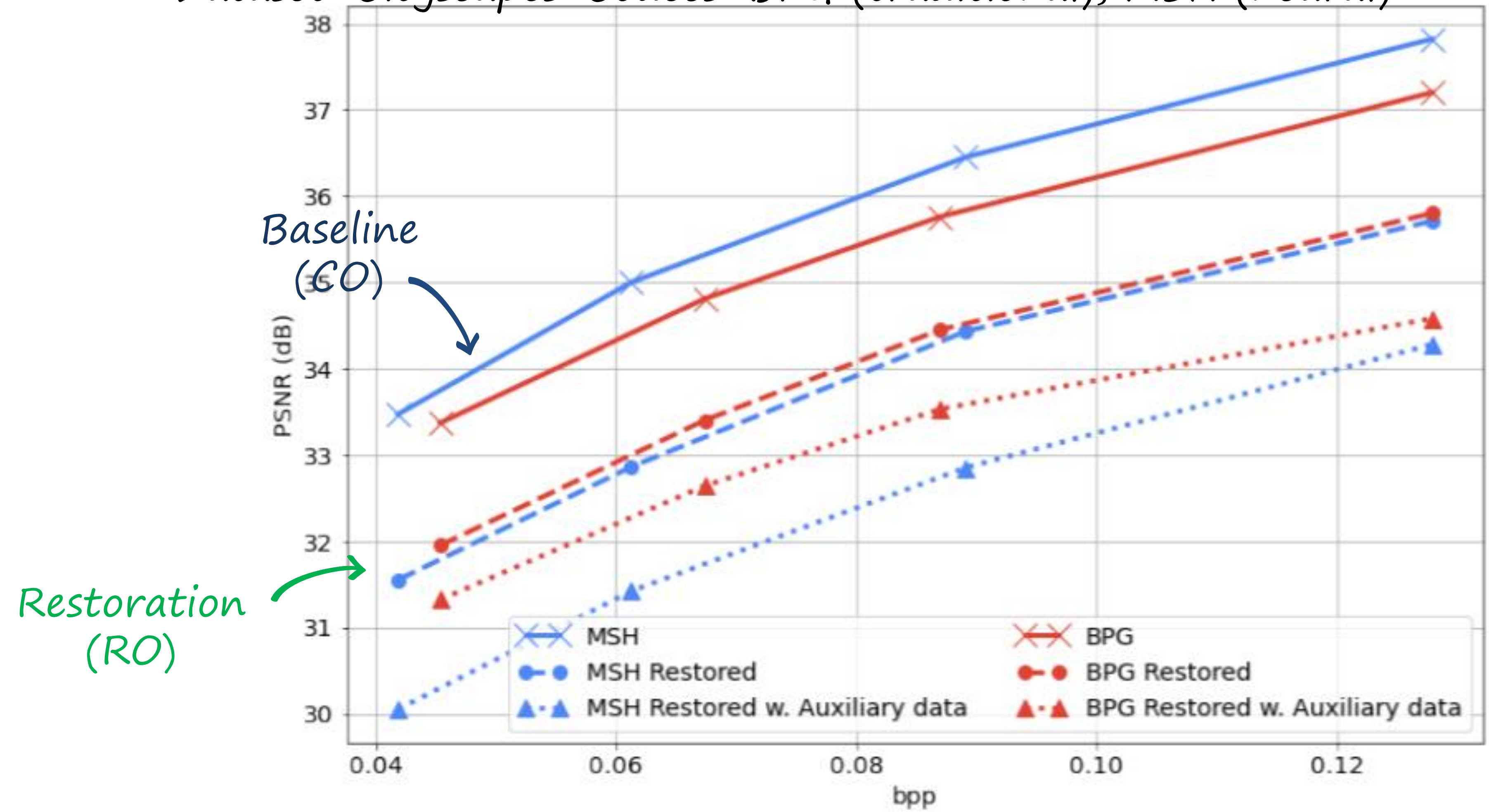


Why does it work?
- Alleviates the covariate shift
- Keeps useful information for segmentation (e.g. texture)

Distributed Learning and Inference with Compressed Images, IEEE Trans. Image Processing 2021

Experiments. Rate-distortion

Dataset: Cityscapes. Codecs: BPG (traditional), MSH (neural)

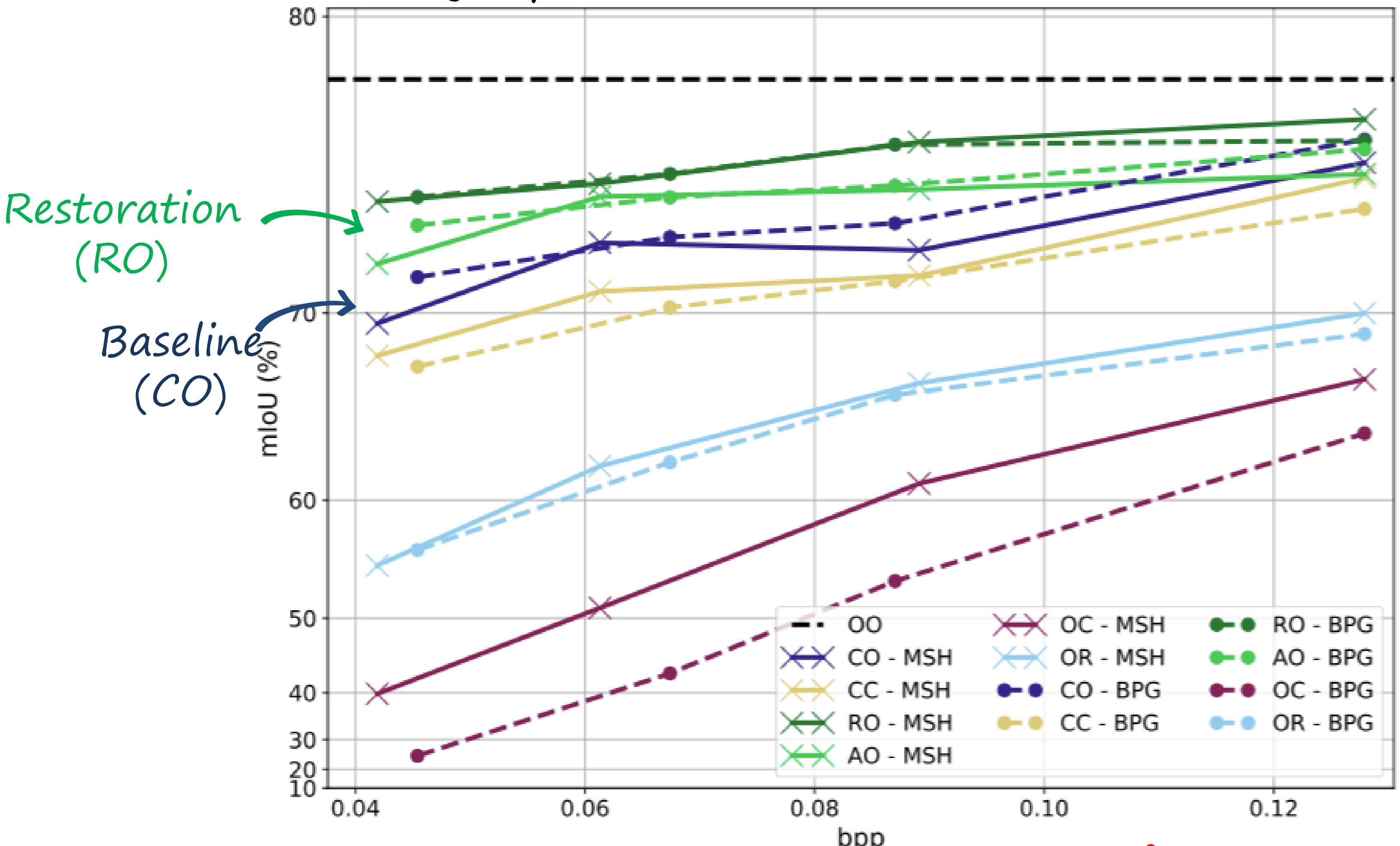


Restoration harms R-D performance

Distributed Learning and Inference with Compressed Images, IEEE Trans. Image Processing 2021

Experiments. Segmentatin

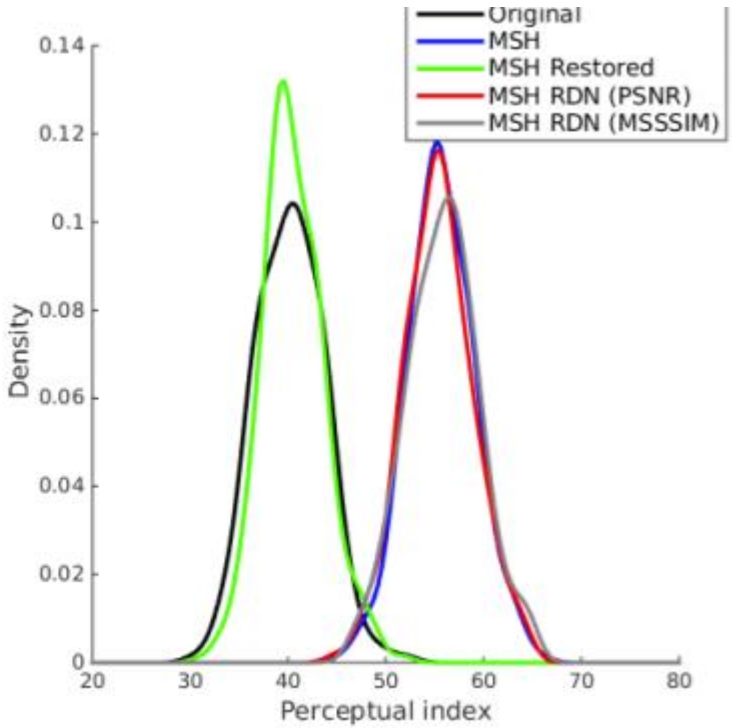
Dataset: Cityscapes. Codecs: BPG (traditional), MSH (neural)

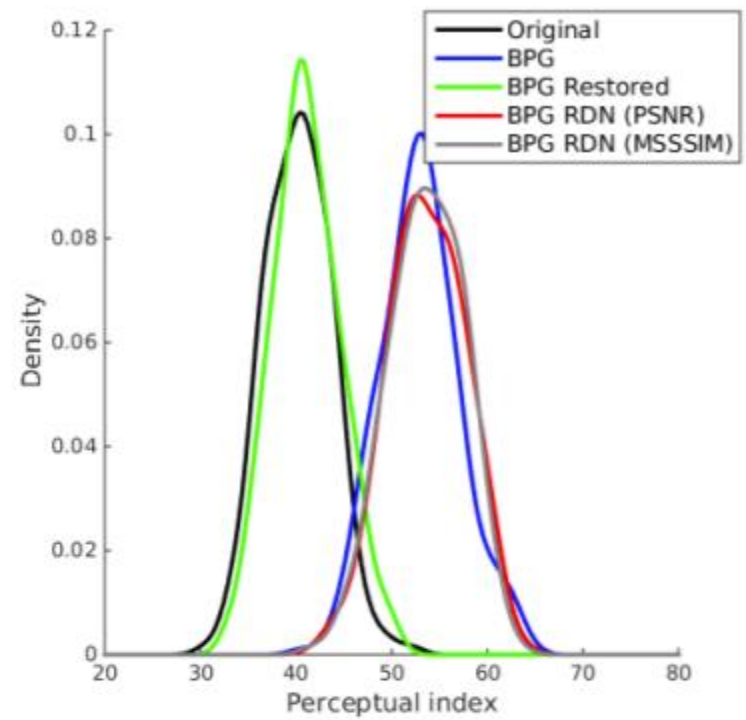


Restoration improves segmentation performance

[Distributed Learning and Inference with Compressed Images](), IEEE Trans. Image Processing 2021

Adversarial vs non-adversarial restoration

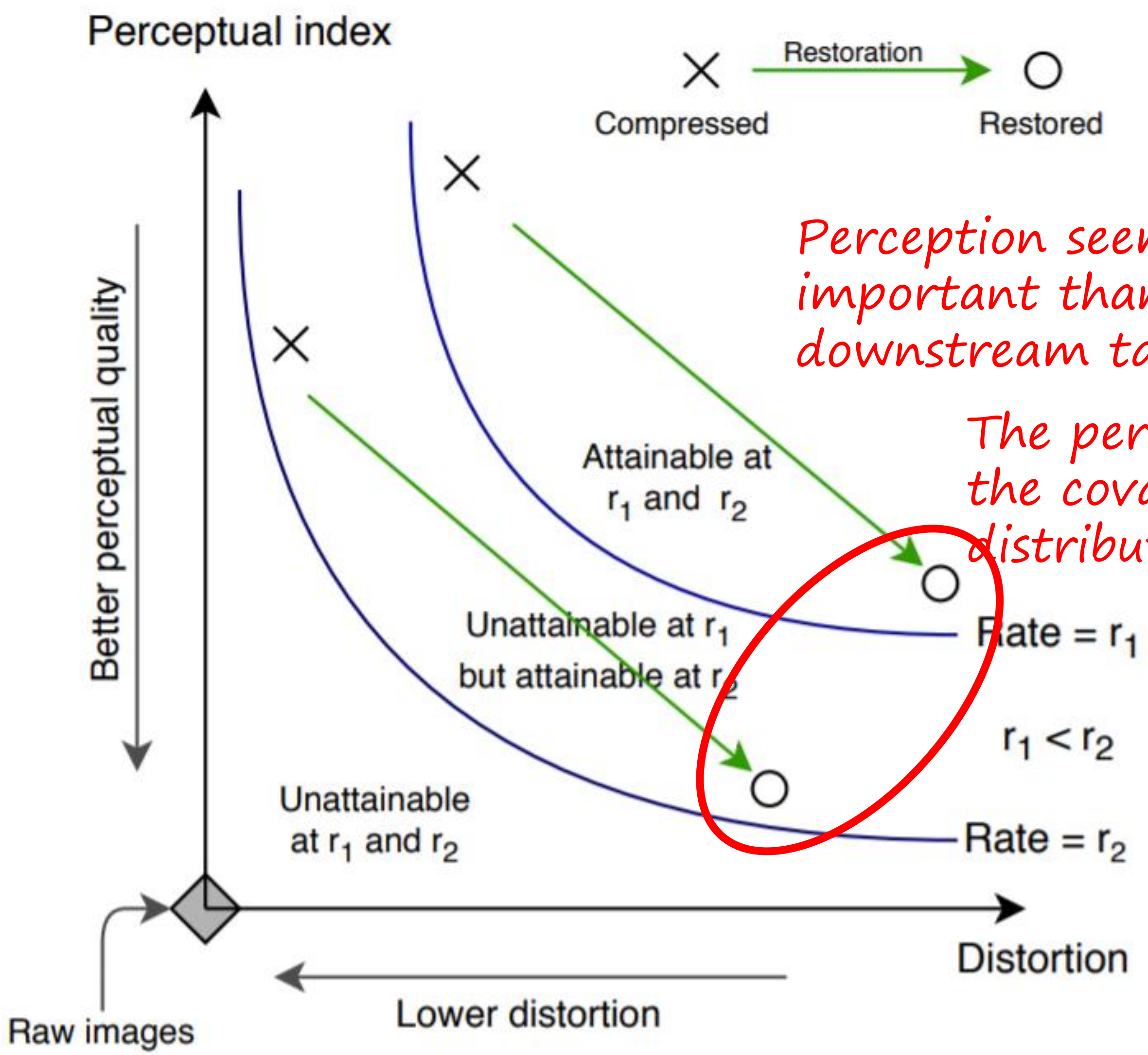




Restoration must be adversarial

[Distributed Learning and Inference with Compressed Images](), IEEE Trans. Image Processing 2021

Perception-distortion tradeoff



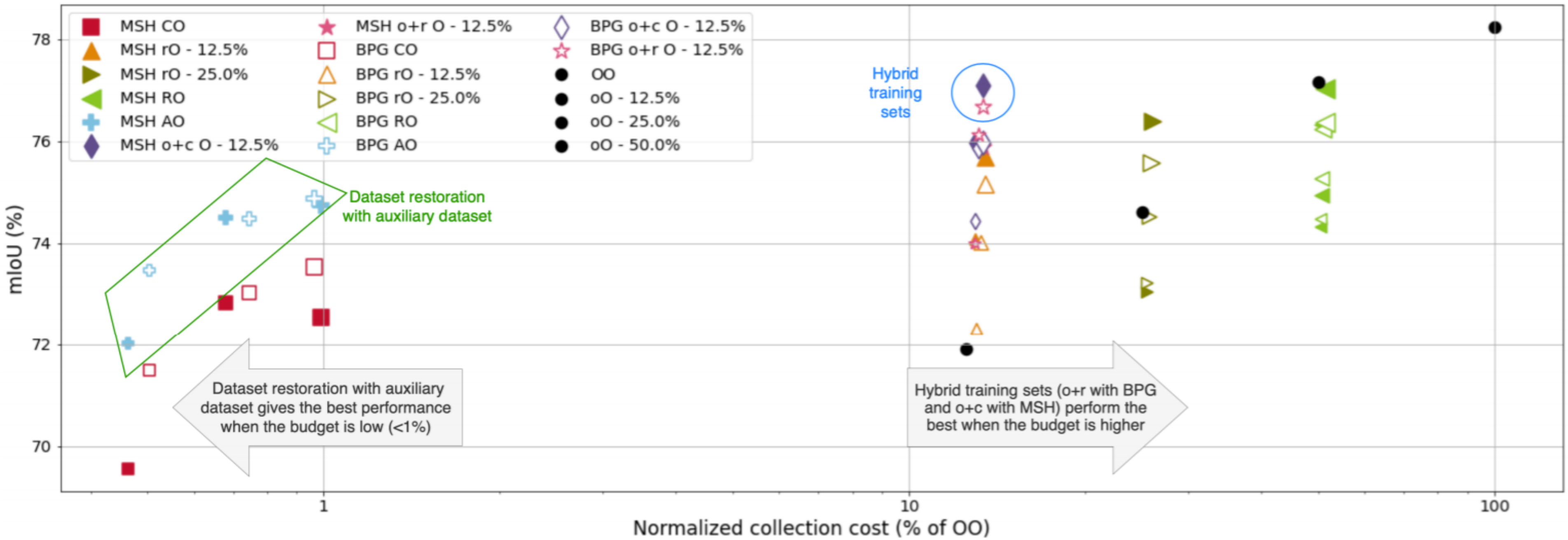
Perception seems to be more important than distortion for downstream tasks

The perceptual index measure the covariate shift wrt the distribution of real images

[Distributed Learning and Inference with Compressed Images](), IEEE Trans. Image Processing 2021

Cost of collecting data

The perceptual index measures
the covariate shift wrt the
distribution of real images



[Distributed Learning and Inference with Compressed Images](), IEEE Trans. Image Processing 2021

References

General references

- Balle et al. End-to-end Optimized Image Compression, ICLR 2017
- Theis et al., Lossy Image Compression with Compressive Autoencoders, ICLR 2017
- Blau and Michaeli, The Perception-Distortion Tradeoff, CVPR 2018
- Blau and Michaeli, Rethinking Lossy Compression: The Rate-Distortion-Perception Tradeoff, ICML 2019
- Dai et al., Deformable Convolutional Networks, ICCV 2017
- Mentzer et al., High-Fidelity Generative Image Compression, NeurIPS 2020

Works by our group and collaborations (with Marta Mrak's group at BBC R&D, London, UK and Shuai Wan's group at Nortwestern Politechnic University, Xi'an, China)

- Yang et al., Variable Rate Deep Image Compression with Modulated Autoencoder, Signal Processing Letters 2020
- Yang et al., Slimmable compressive autoencoders for practical imaga compression, CVPR 2021
- Katakol et al., DANICE: Domain adaptation without forgetting in neural image compression, CLIC 2021 at CVPR 2021
- Zhang et al., DCNGAN: A deformable convolution-based GAN with QP adaptation for perceptual quality enhancement of compressed video, ICASSP 2022
- Katakol et al., Distributed Learning and Inference with Compressed Images, IEEE Trans. Image Processing 2021

THANK YOU!

lherranz@cvc.uab.es
www.lherranz.org

CVC
UAB Universitat Autònoma de Barcelona
MINISTERIO DE CIENCIA E INNOVACIÓN
UNIÓN EUROPEA
FONDO SOCIAL EUROPEO
El FSE invierte en tu futuro
AGENCIA ESTATAL DE INVESTIGACIÓN