ETH zürich

Organisationseinheit verbal
optional auf 2 Zeilen

Controlling Rate, Distortion, and Realism: Towards a Single Comprehensive Neural Image Compression Model

Zixuan Chen

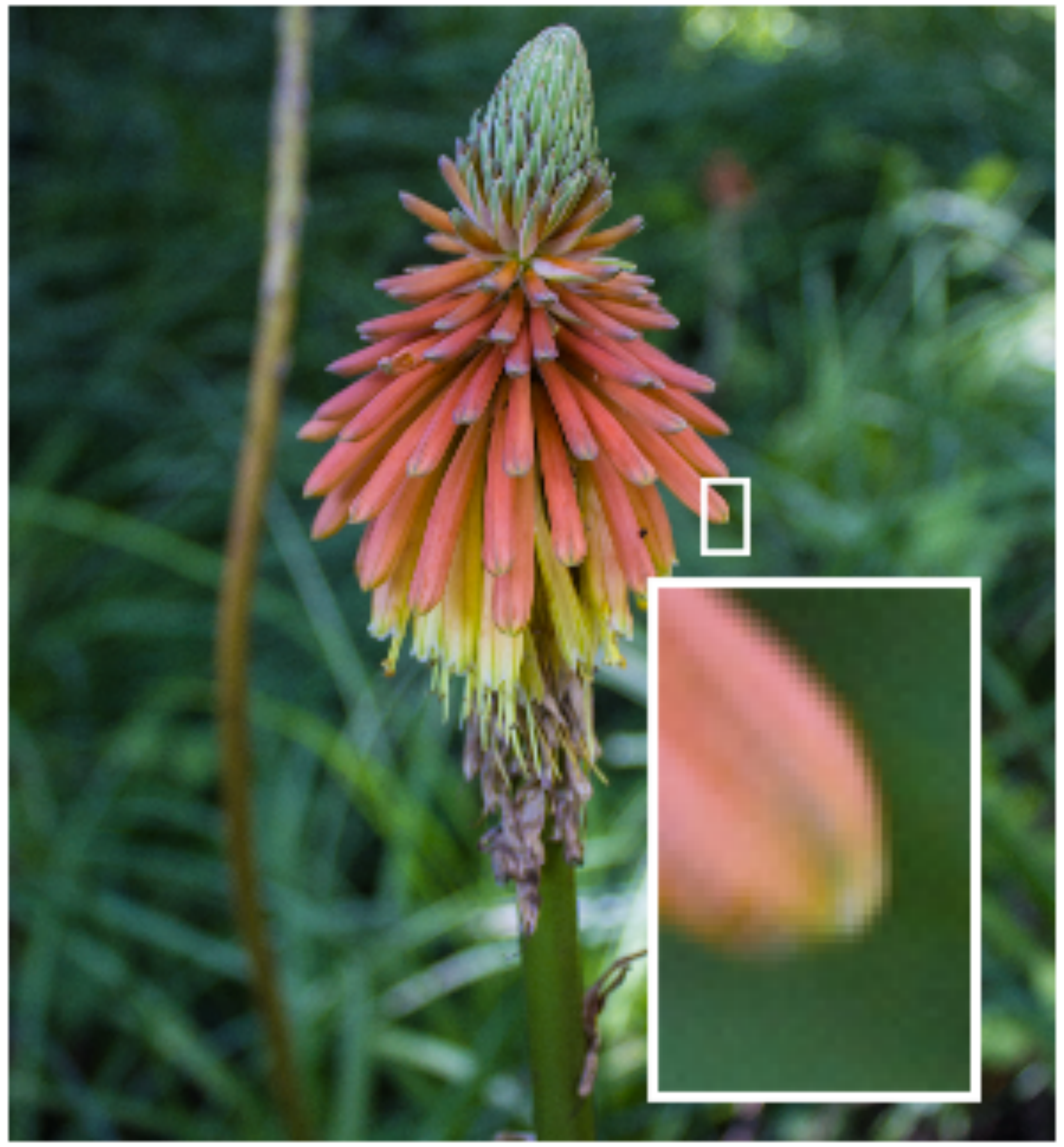
Mentor: Till Aczel

Seminar in Deep Neural Networks

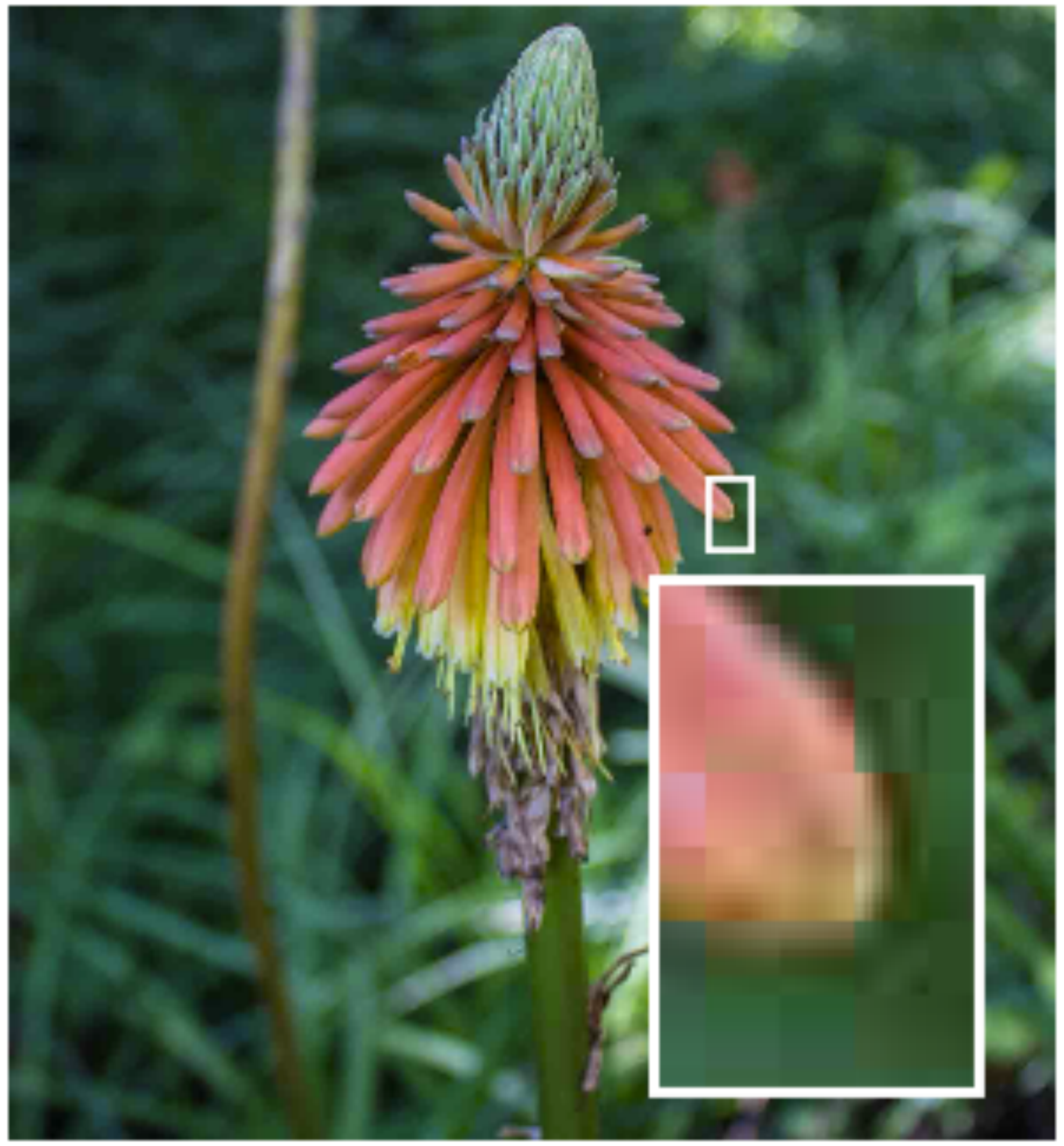
26.03.2024, ETH Zurich

Introduction: Image Compression

Usage: image storage and transmission

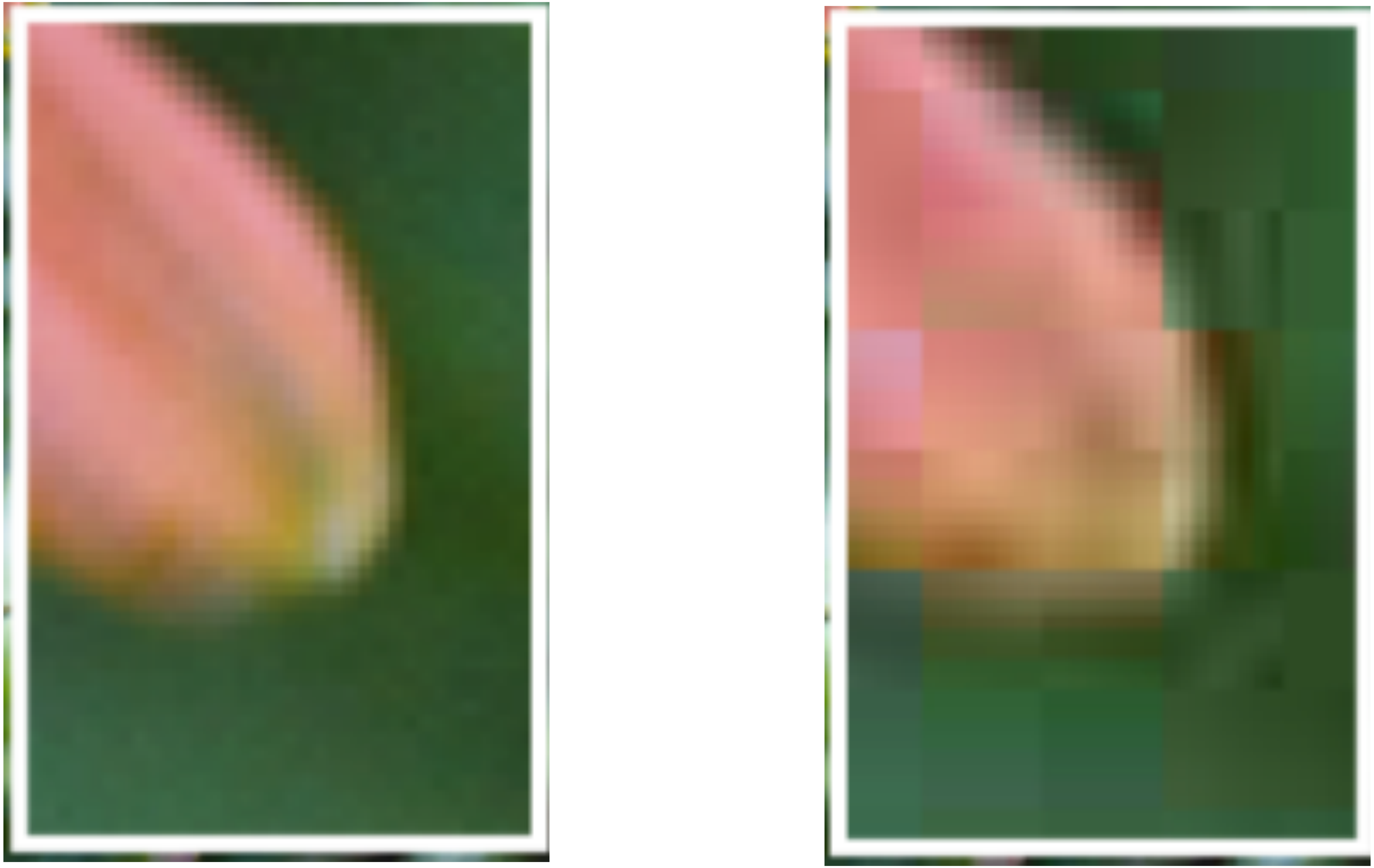


8.9M



68.34K

From https://helpx.adobe.com/au/lightroom-classic/lightroom-key-concepts/compression.html



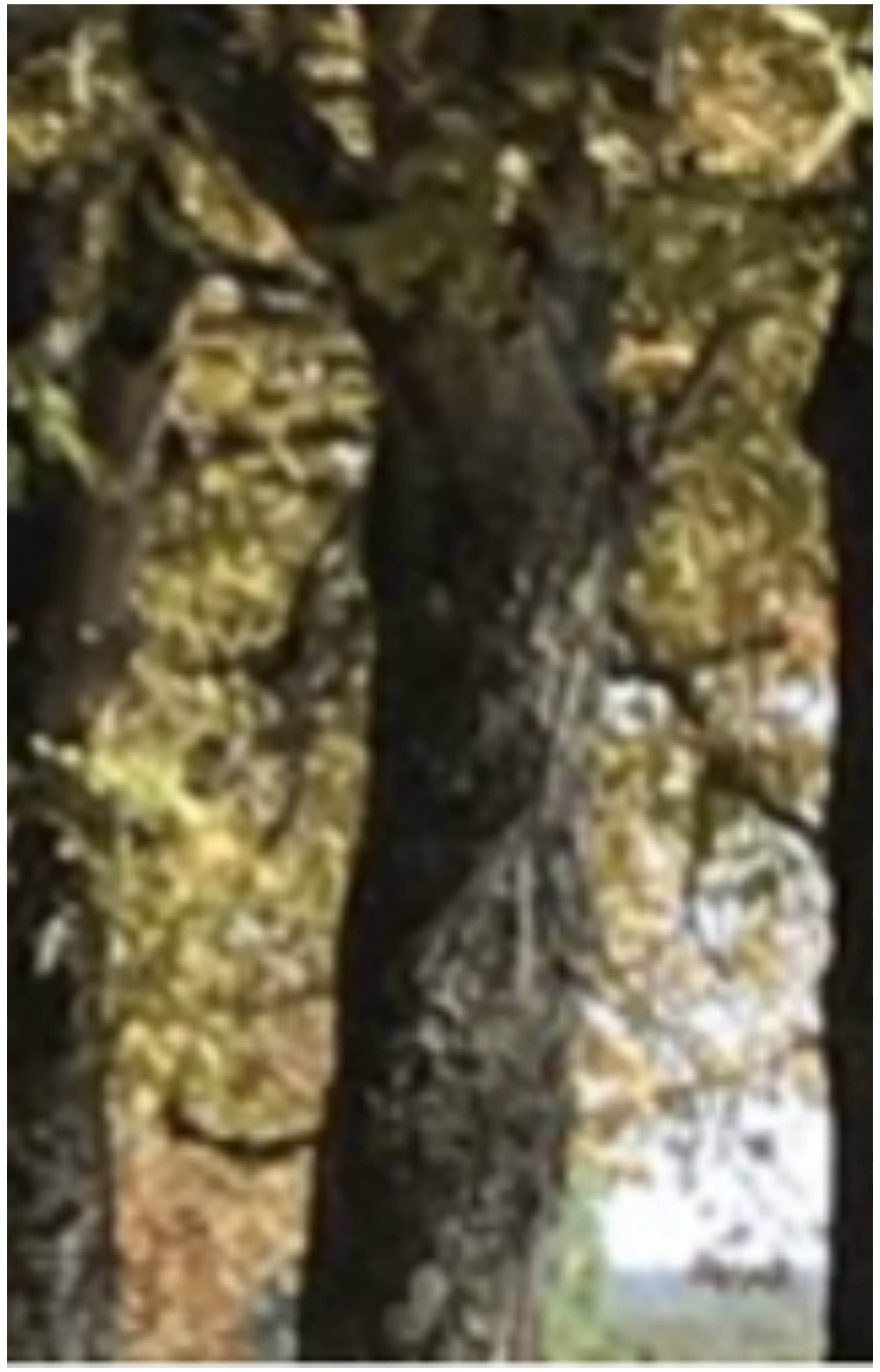
From https://helpx.adobe.com/au/lightroom-classic/lightroom-key-concepts/compression.html

Three main indicators:
Rate, distortion, and realism

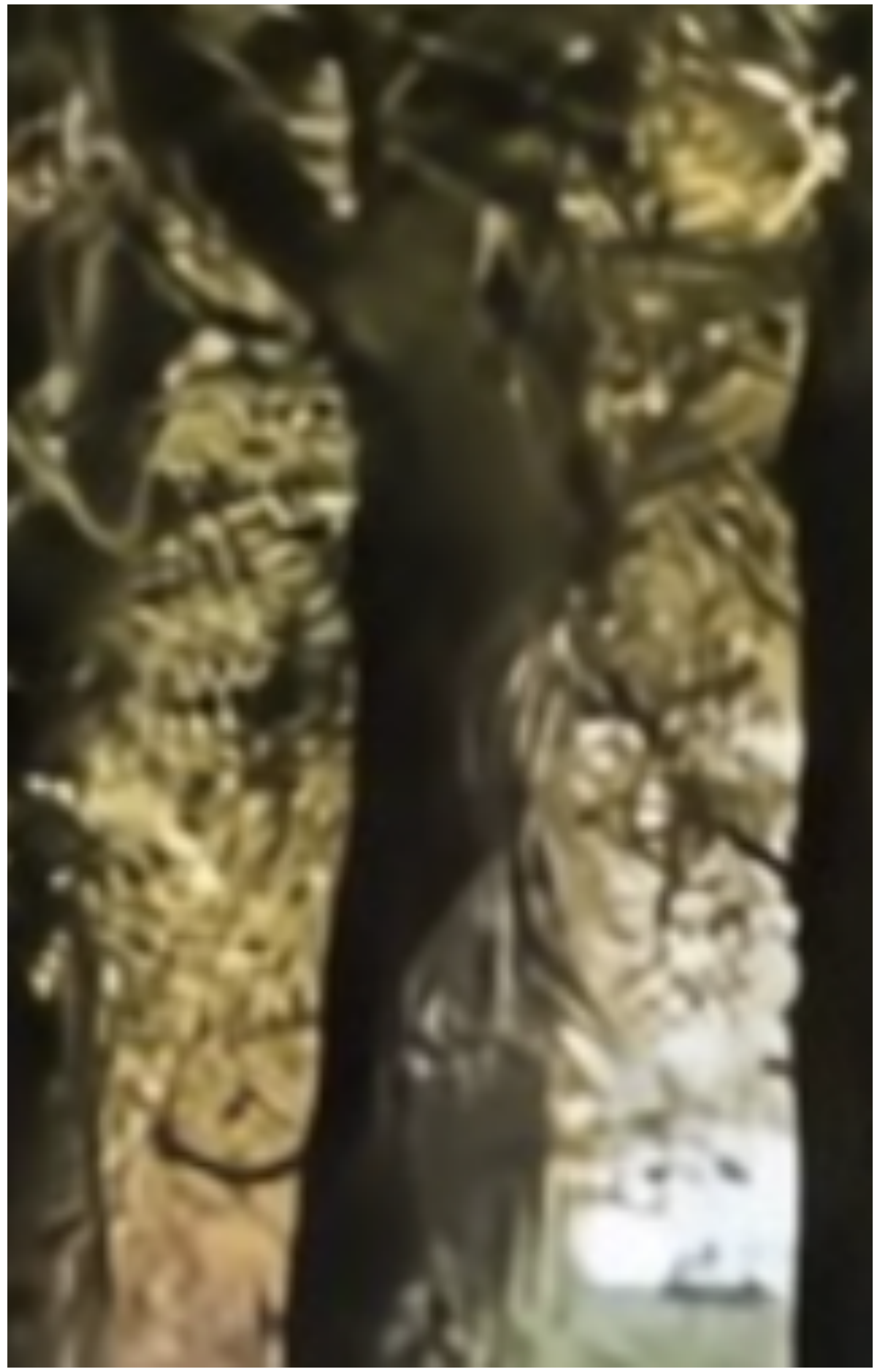
Guess and discuss

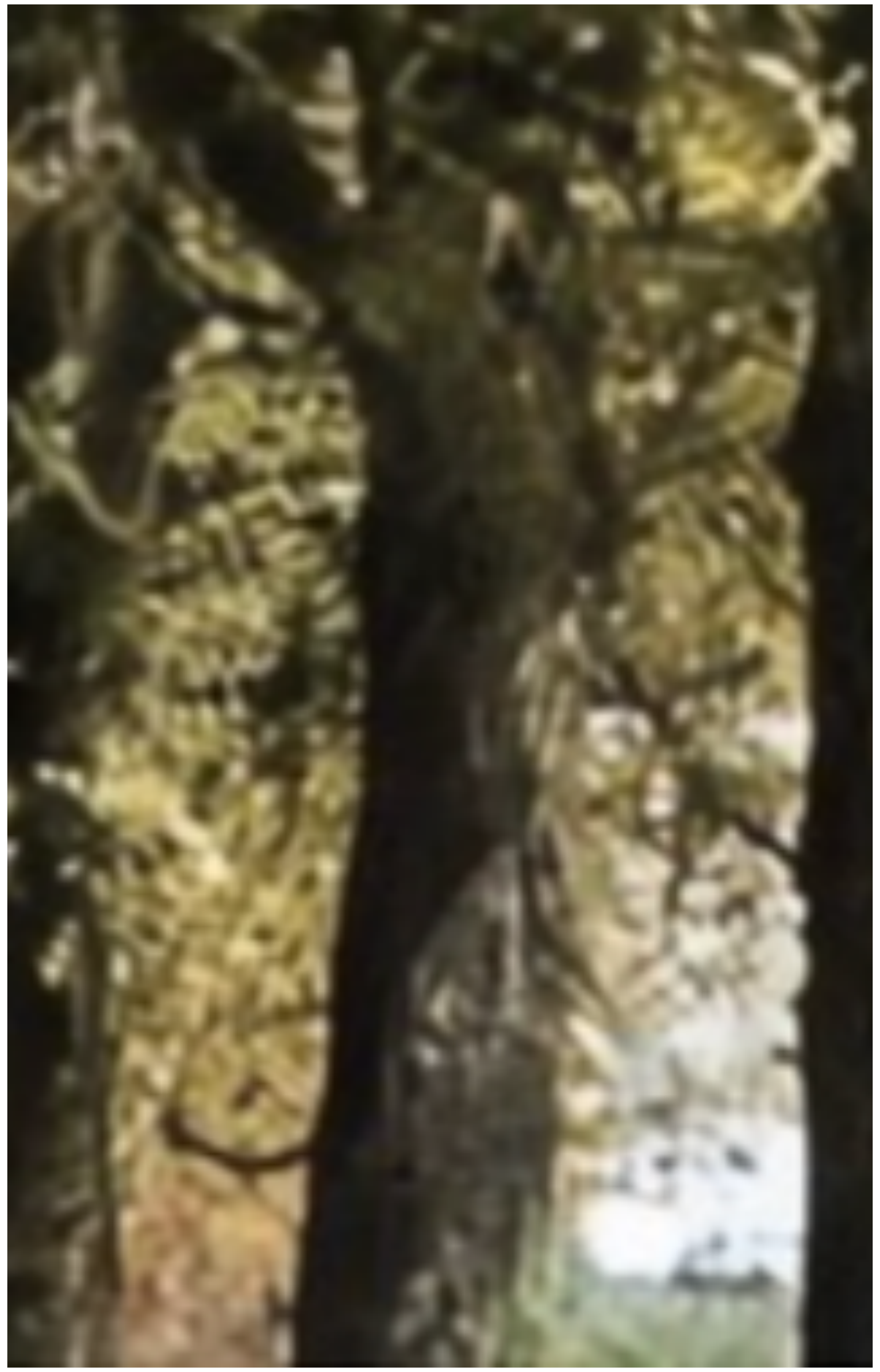
Which two are of the same bit rate?

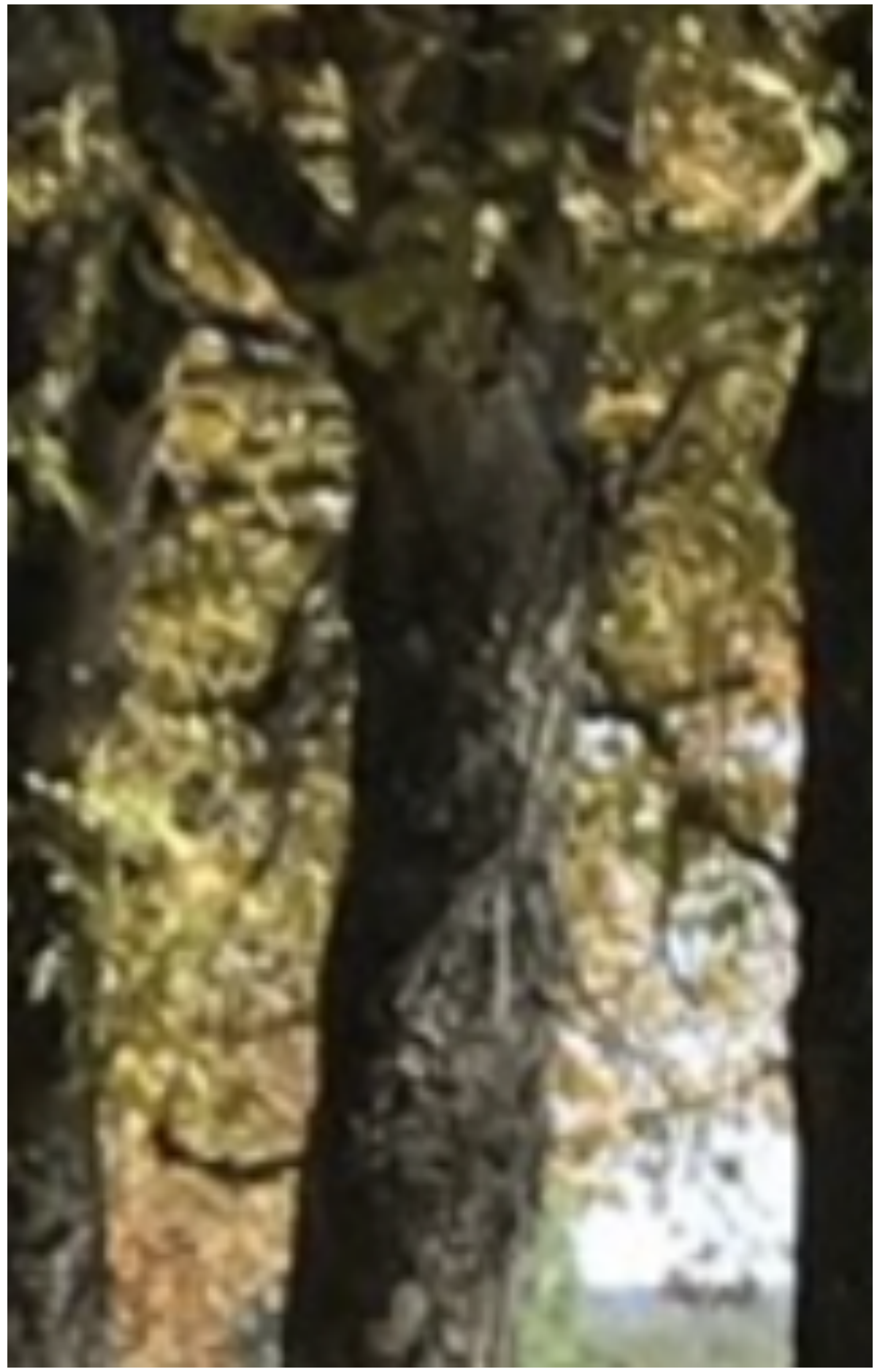
Which one maintains the most details?



original



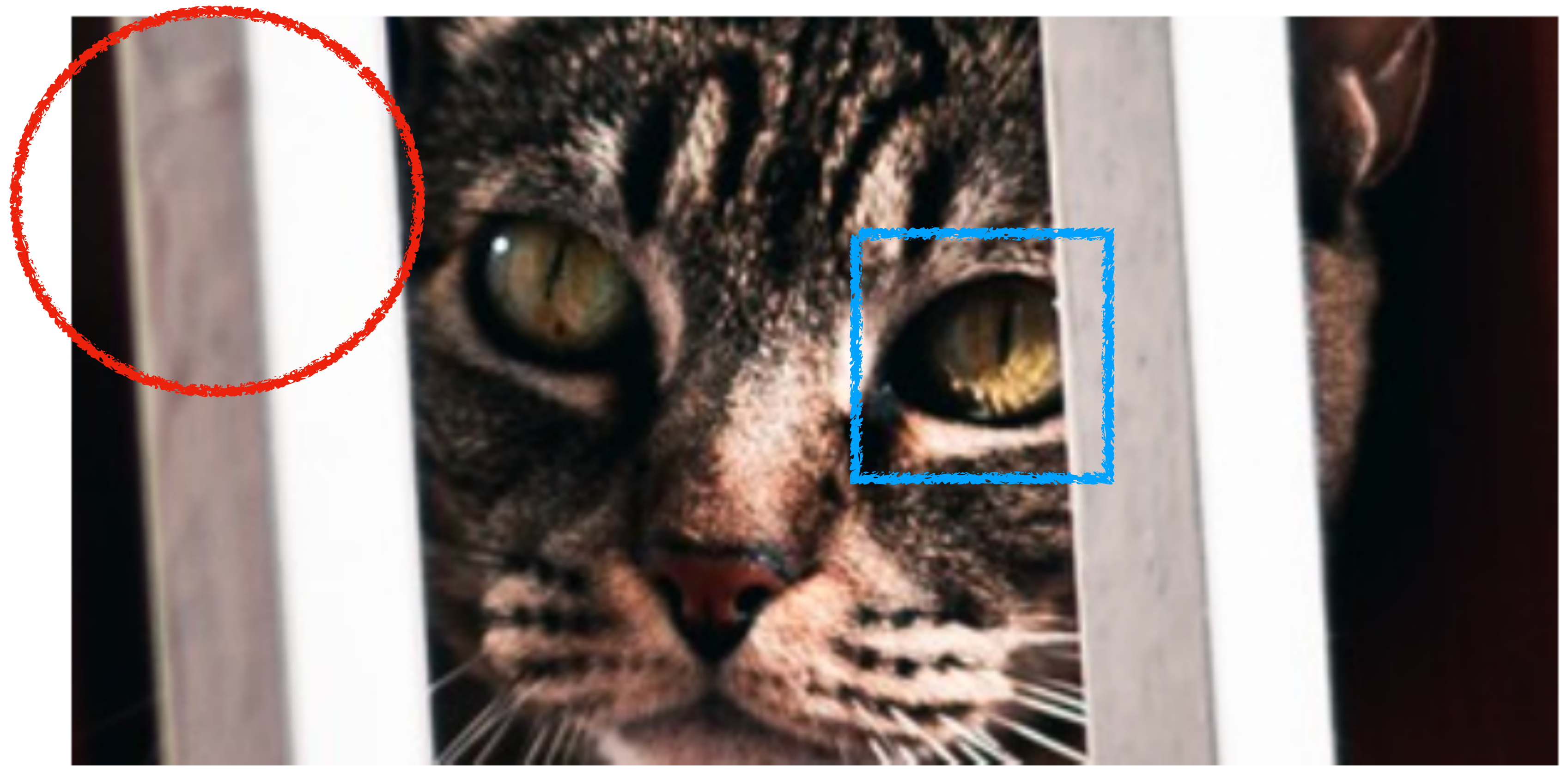


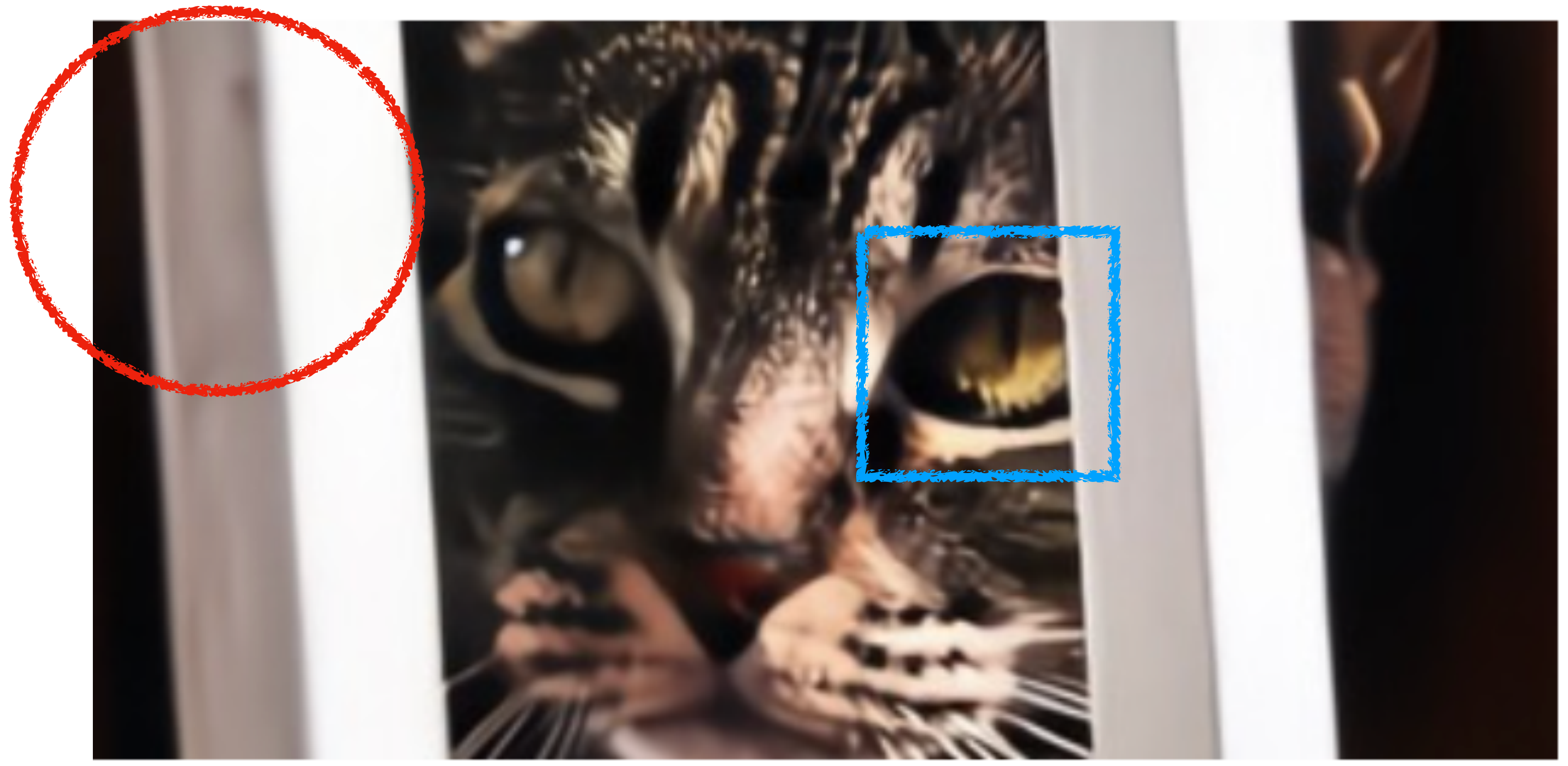


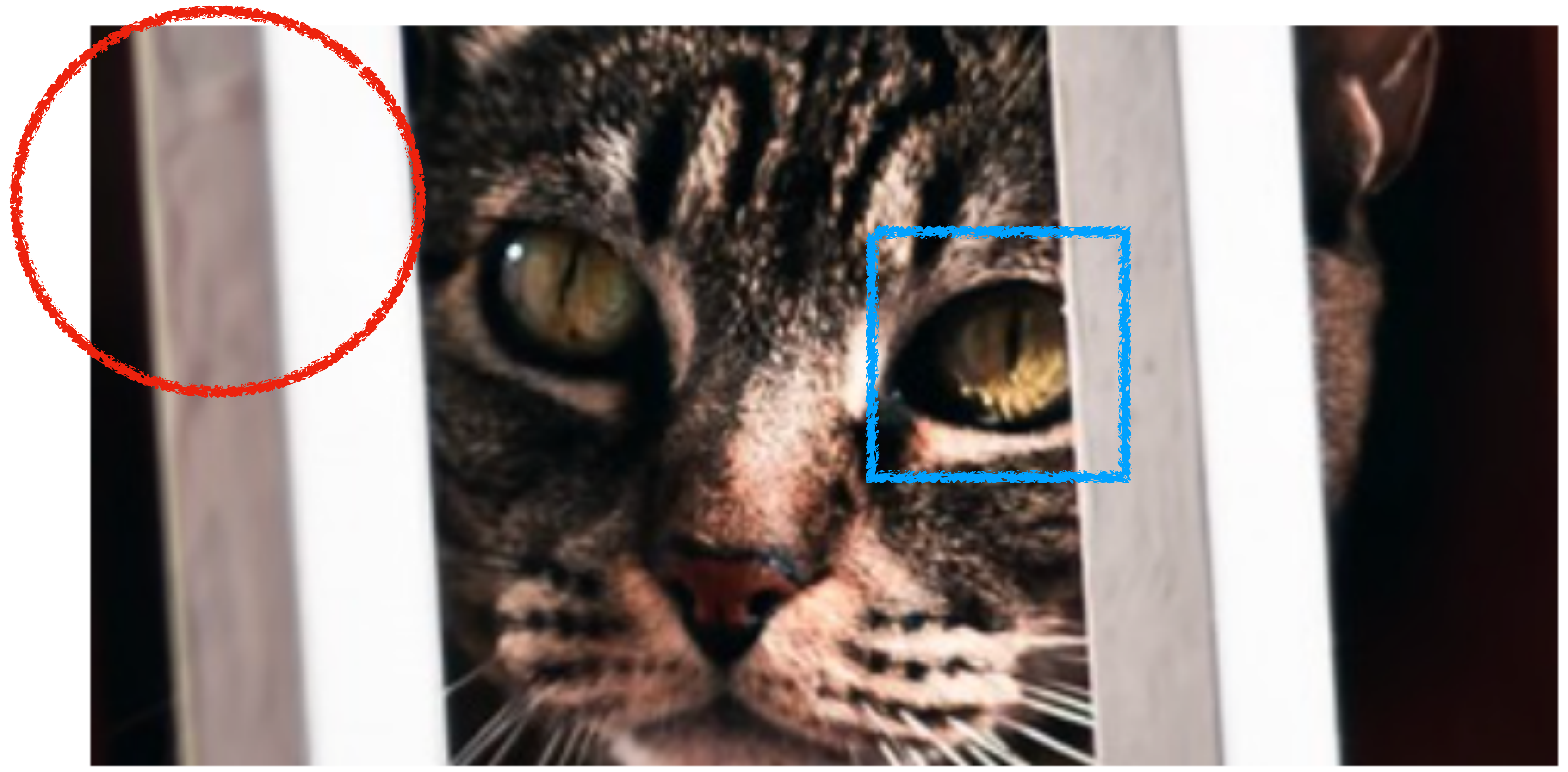
From Iwai et al. [*thisp]

Bit-rate: BPP(bits per pixel)

original

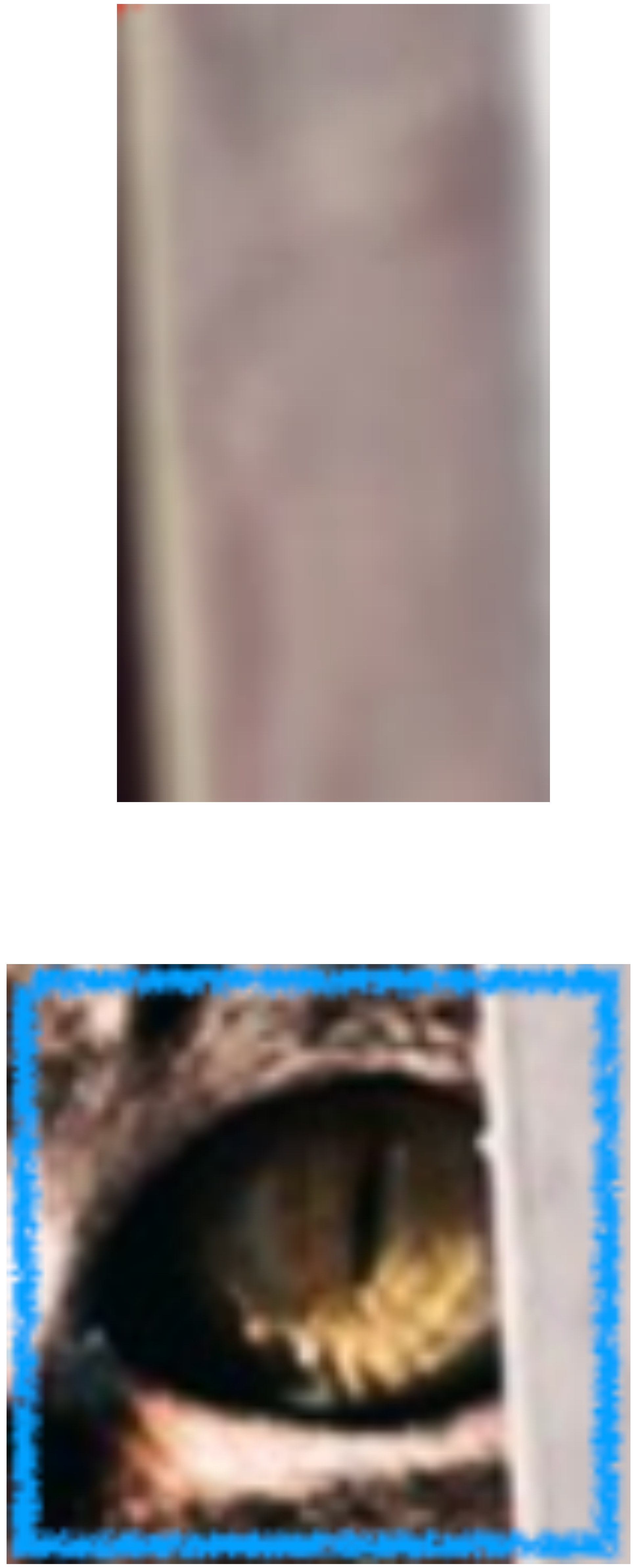




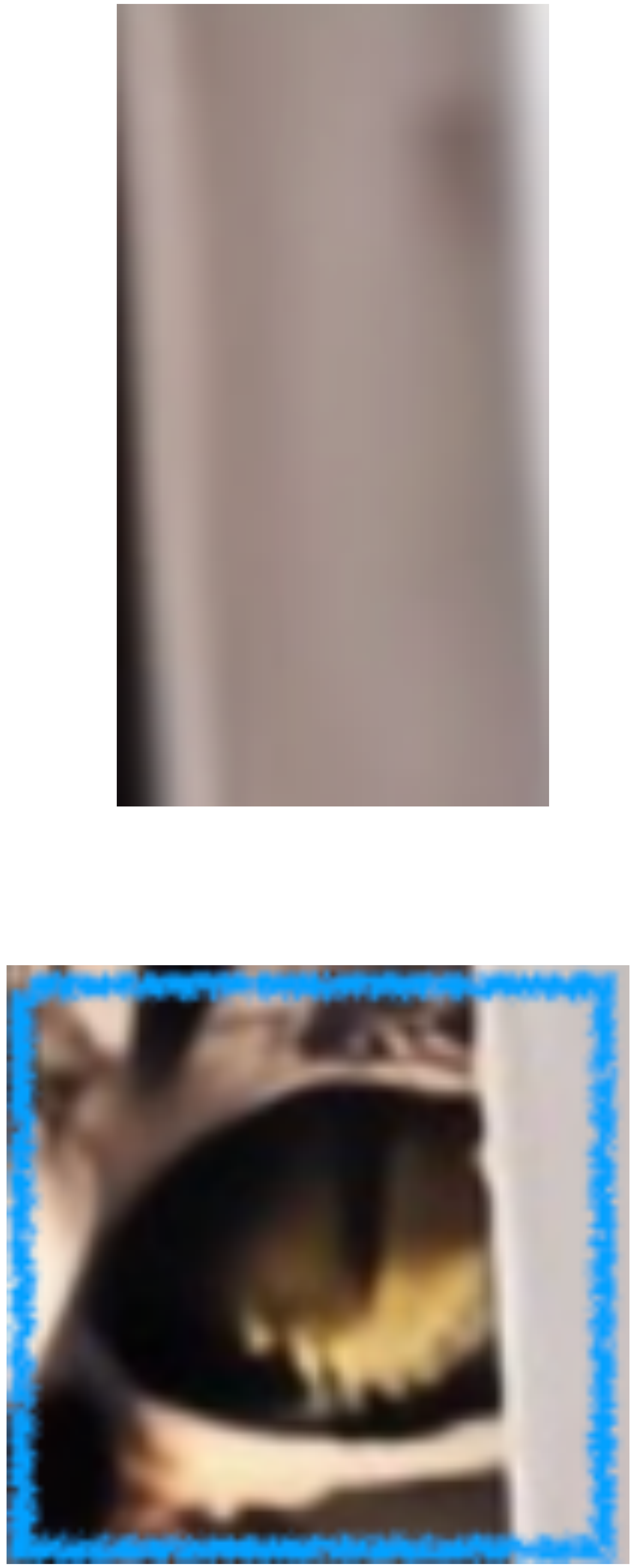


From Iwai et al. [*thisp]

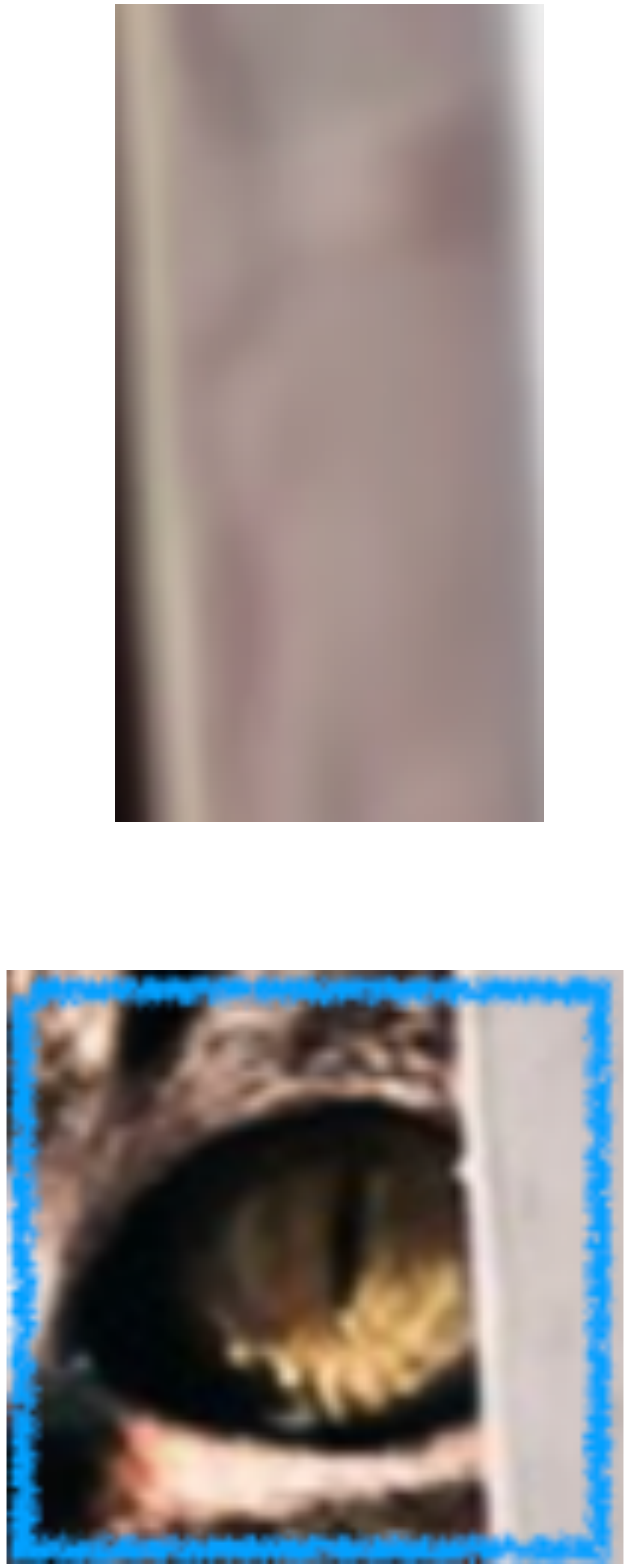
Bit-rate



origin



low rate



high rate

From Iwai et al. [*thisp]

Measurements of Distortion

$$\begin{aligned} MSE(f,g) &:= \frac{1}{|V|}\int_V |f(x)-g(x)|^2\ dx \quad \textit{(Mean Squared Error)}, \\ PSNR(f,g) &:= 10\log_{10}\left(\frac{m^2}{MSE(f,g)}\right) \quad \textit{(Peak Signal to Noise Ratio)} \end{aligned}$$

V: a rectangular region of the image

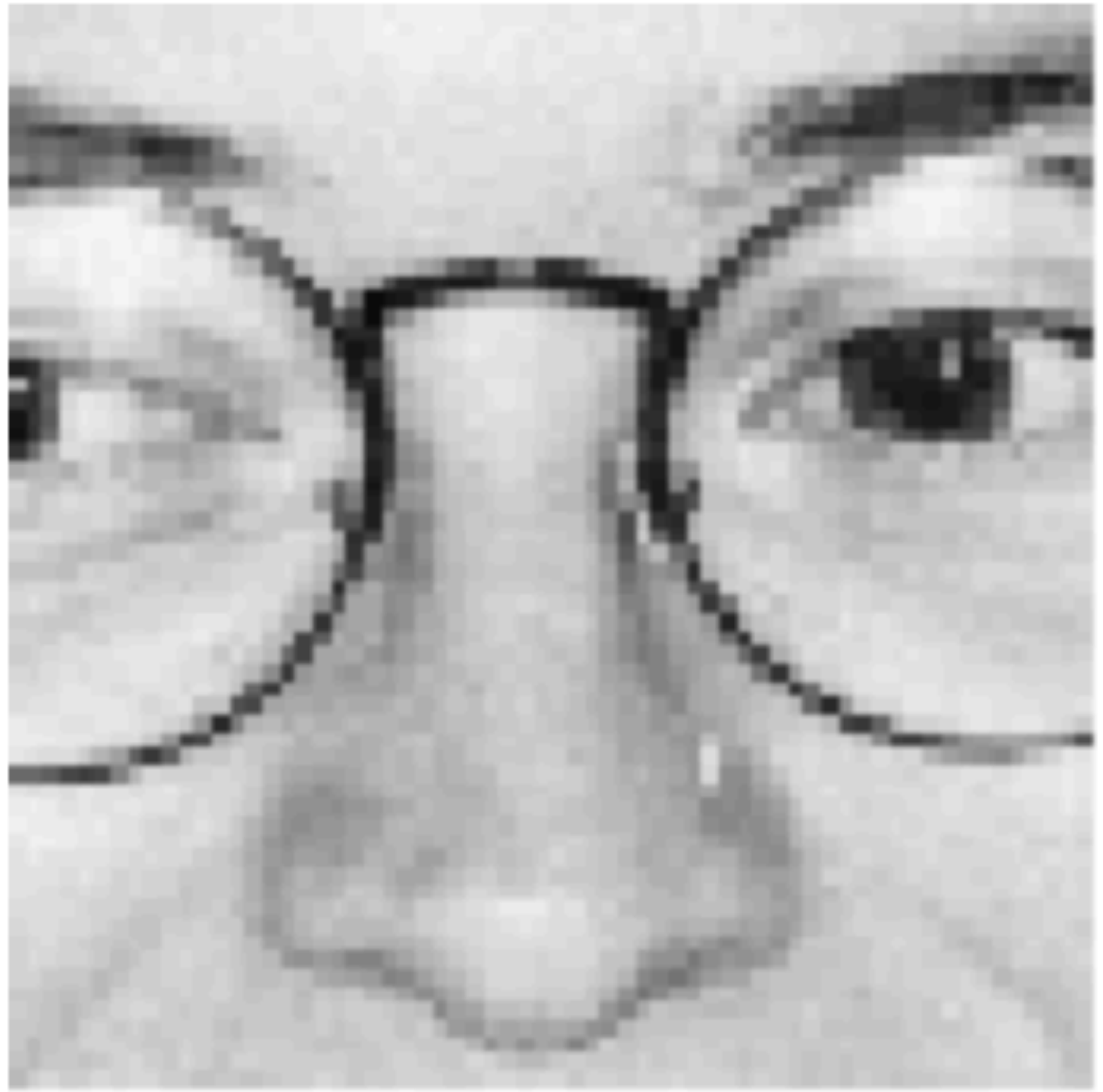
f, g: images

m: the maximum possible pixel value of the image

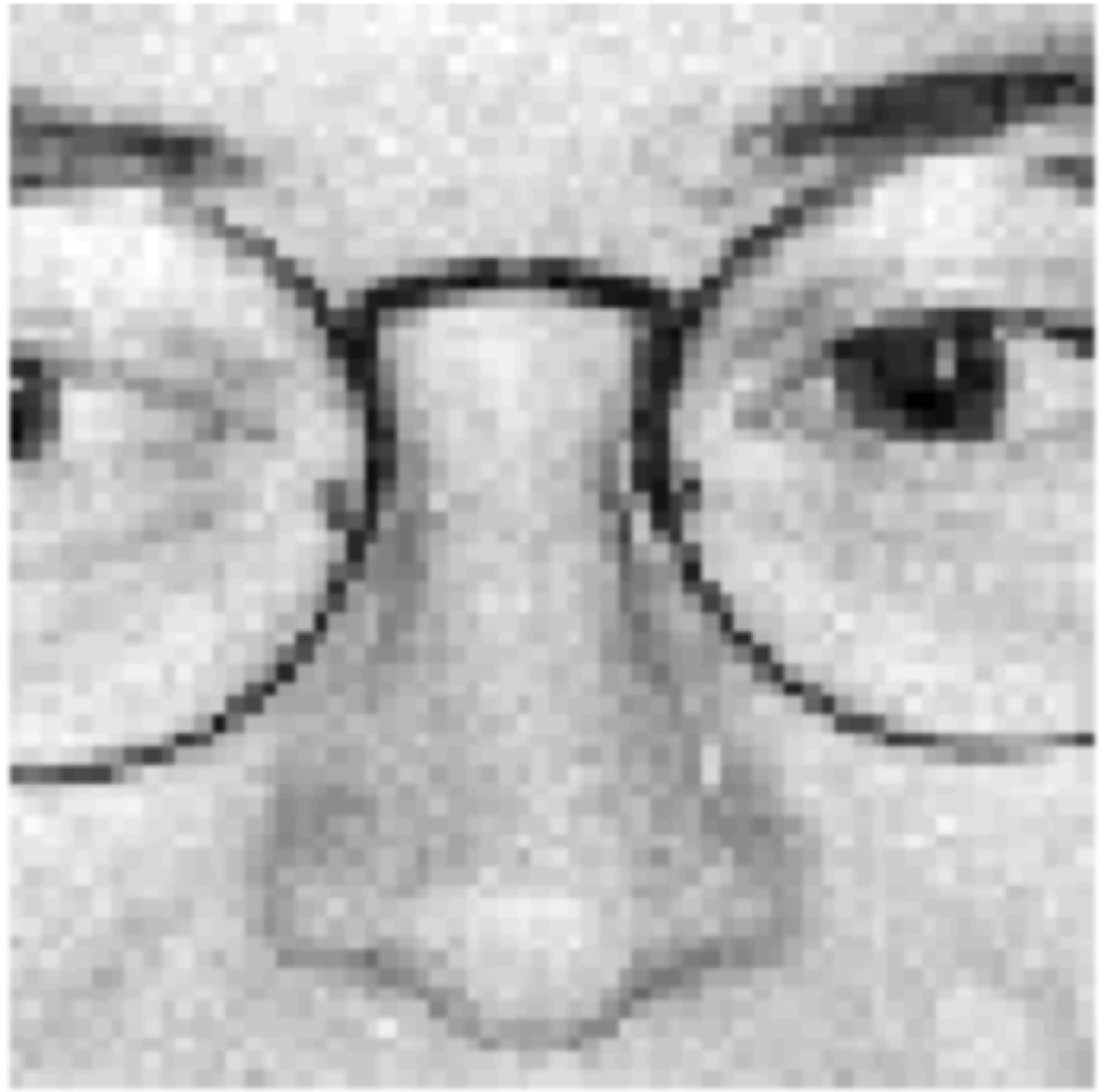
From Becker Axel [*measure]

Distortion

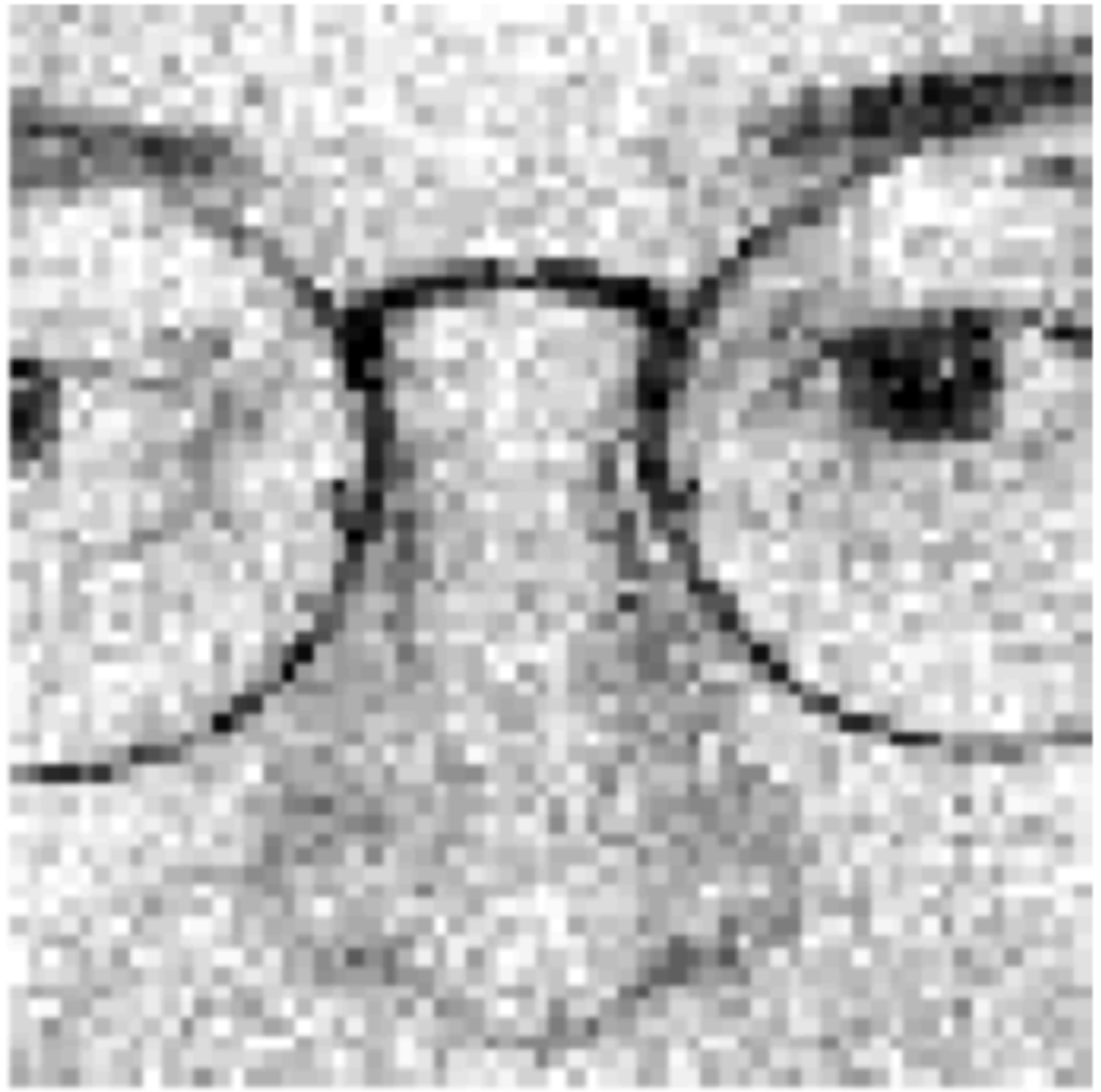
lower distortion = closer to the original image



PSNR = 40 dB



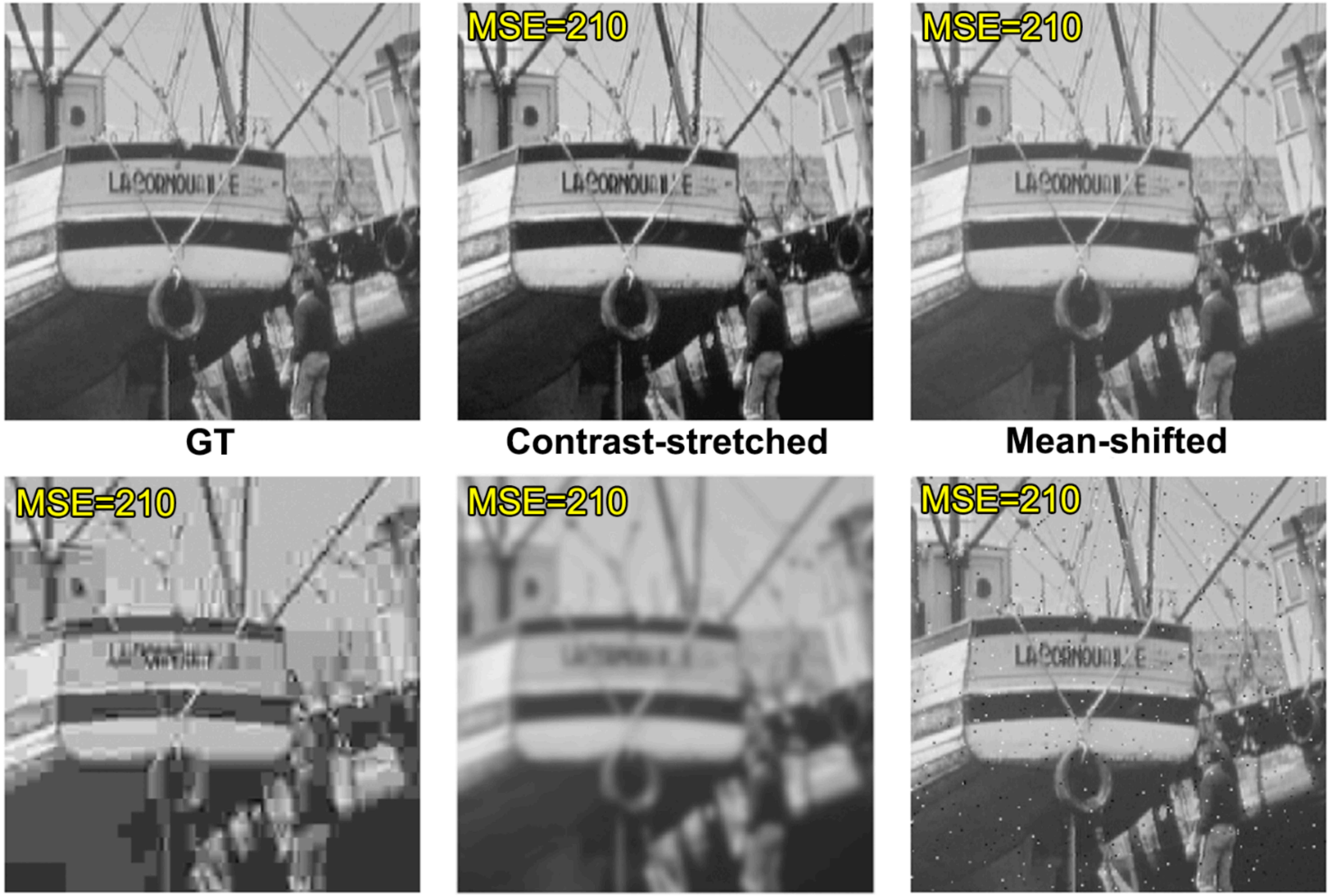
PSNR = 30 dB



PSNR = 20 dB

From https://homepages.inf.ed.ac.uk/rbf/CVonline/LOCAL_COPIES/VELDHUIZEN/node18.html

Realism ≠ distortion



GT

Contrast-stretched

Mean-shifted

JPEG-compressed

Blurred

Salt-pepper noise

From Wang, et a. [*sameMSE]

Measurements of Realism

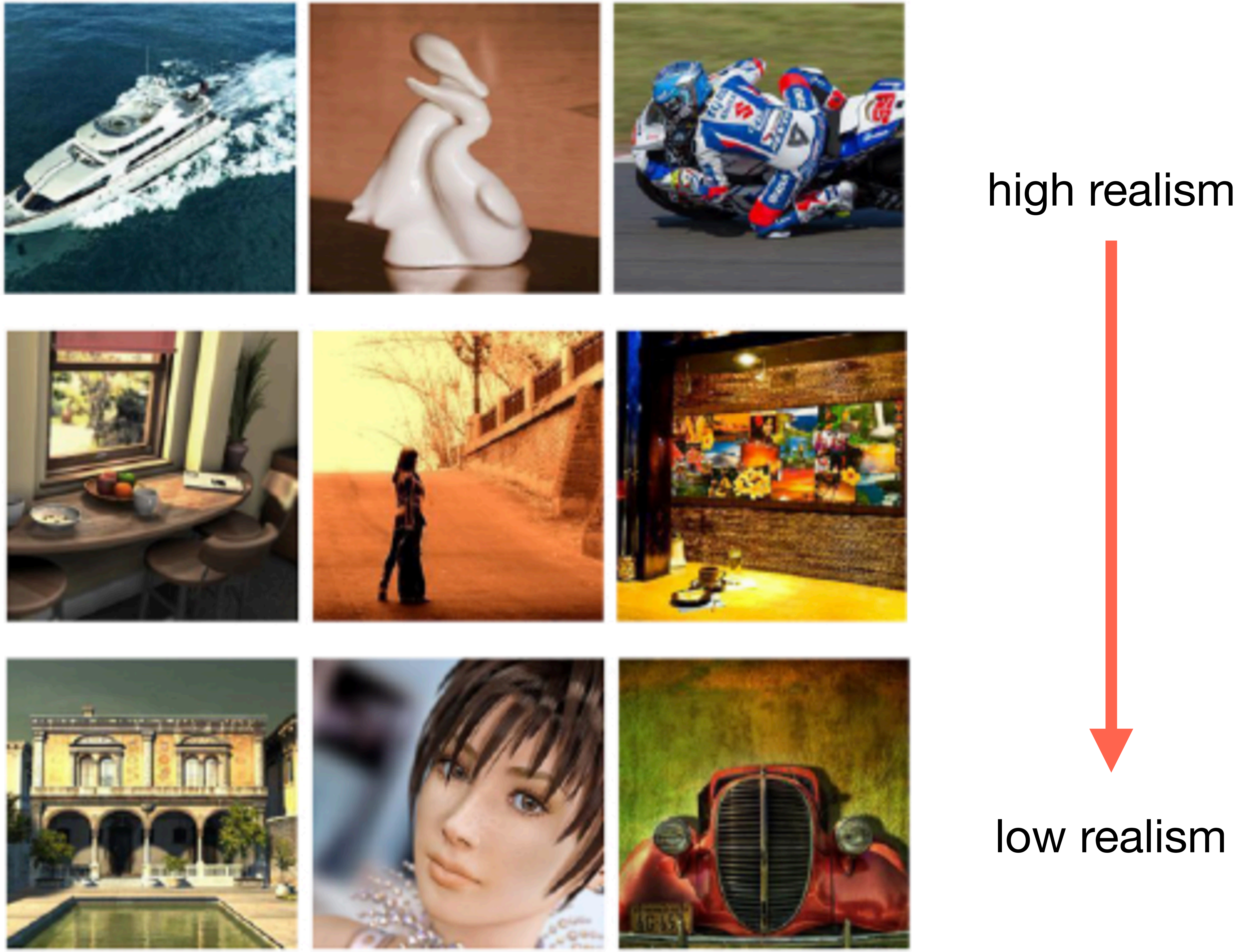
1. FID
2. LPIPS

- Both use deep convolutional networks.

Realism

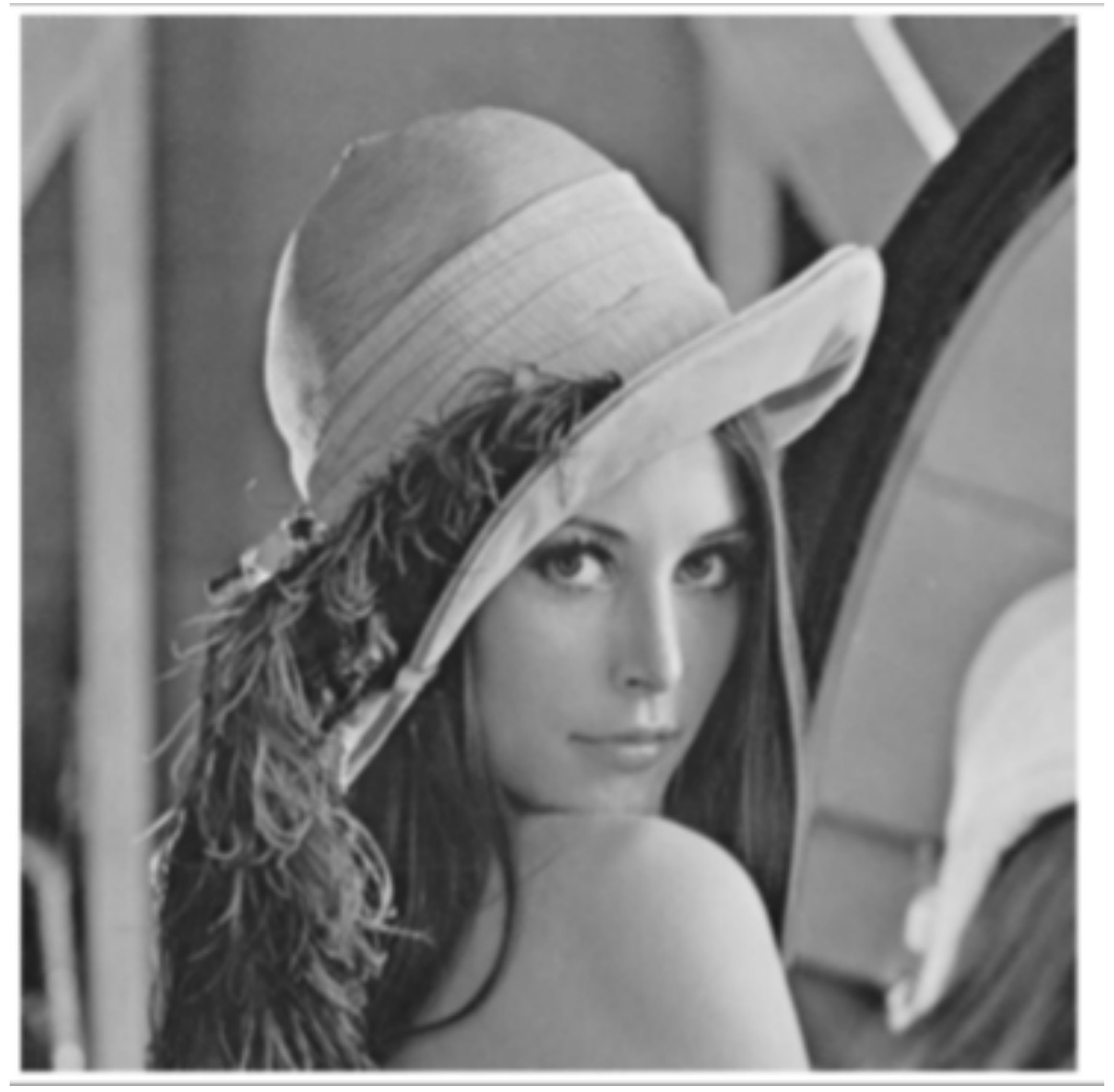
low FID

~ high realism



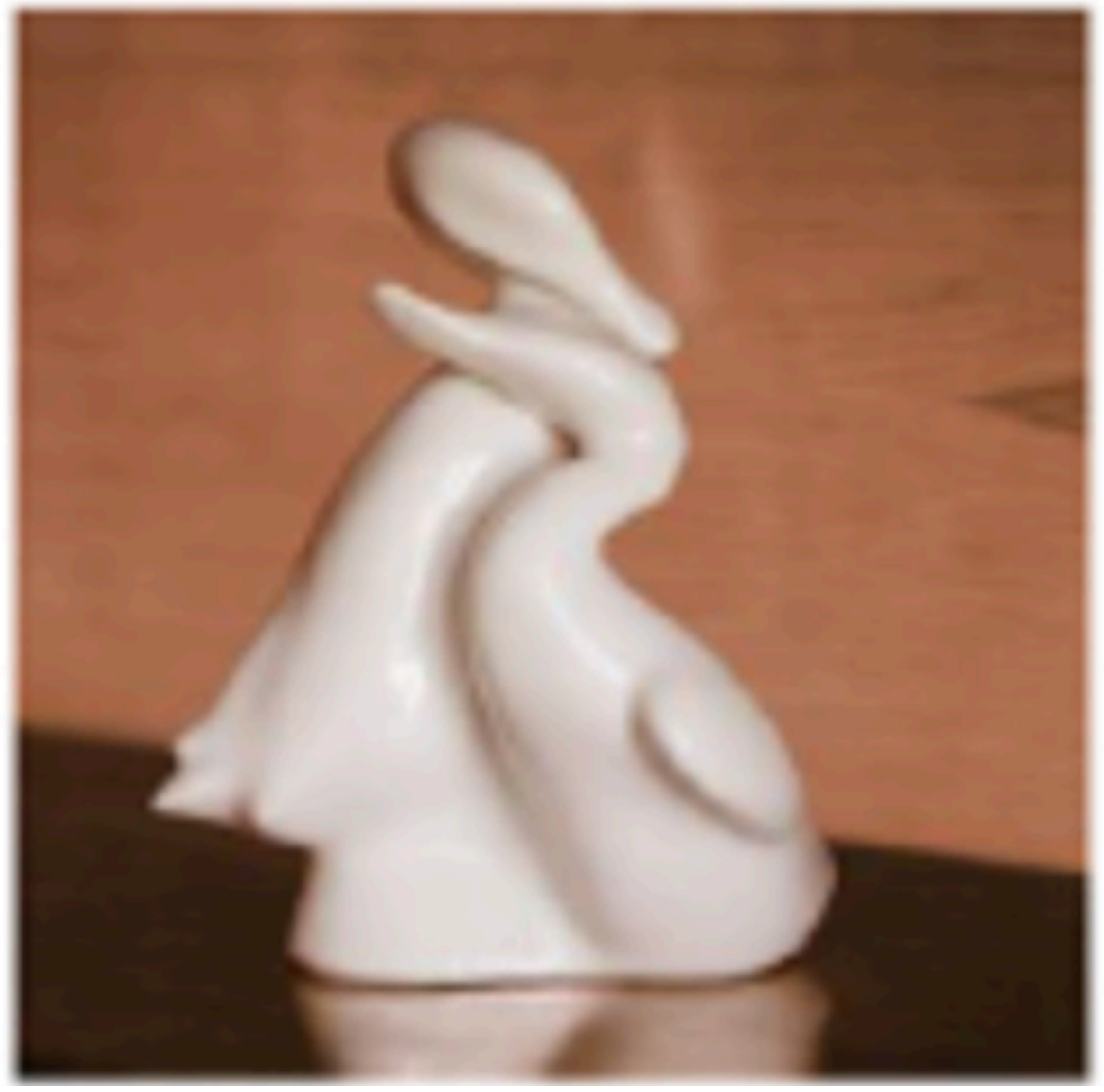
From Fan et al. [*realism]

High Realism itself makes no sense



original

→



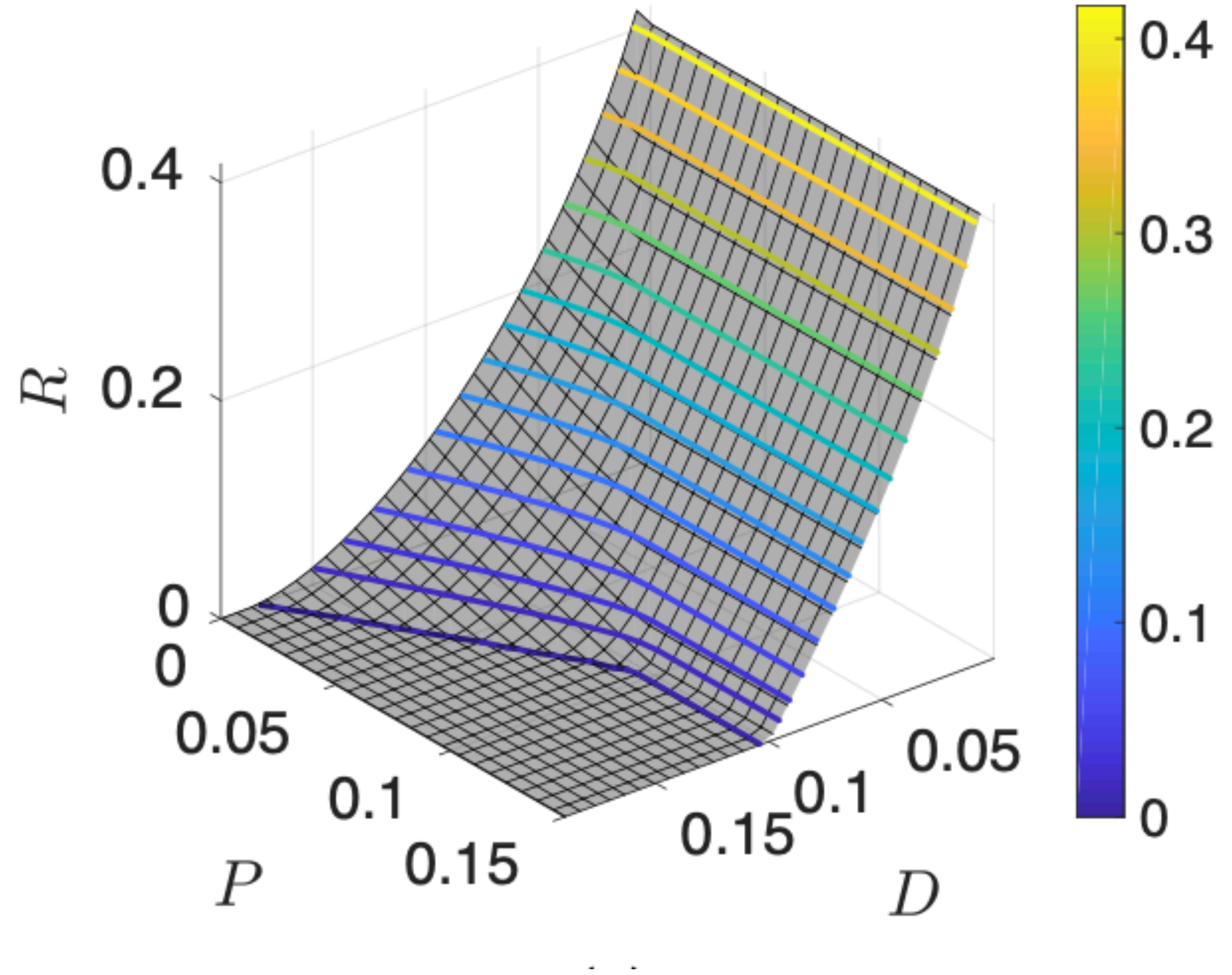
reconstructed

We want low bit rate, low distortion, and high realism!

However, these three indicators cannot be achieved simultaneously!

Rate-distortion-realism Tradeoff

(Curves on the blackboard)



- Fix rate R
- Fix distortion D
- Fix FID’s upper bound P

From Yochai Blau and Tomer Michaeli [*tradeoff]

Tradeoff: takeaway messages

- At low bit rates, the tradeoff becomes stronger.
- To optimize one metric, the other two need to be sacrificed.

From Yochai Blau and Tomer Michaeli [*tradeoff]

A Classic Compression Pipeline:

Single-rate, no realism control

Single-rate v.s. Variable-rate

Loss function

Previous Work

- Learning based:
    - Generative:
        - GAN-based: Multi-realism, HiFiC, PQMIM
        - Diffusion-based: HFD, DIRAC
    - Non-Generative: ELIC, Charm, IVR, Hyperprior
- Non-Learning based: VTM, JPEG

“Green”: able to adjust Distortion-realism tradeoff in one model

GAN Based Training

Motivation: how to adjust the balance between rate, distortion, and realism within a single model?

This Paper: Pipeline

(See Blackboard)

beta: realism weight (higher beta means higher realism and higher distortion, vice versa)

This Paper: Loss function

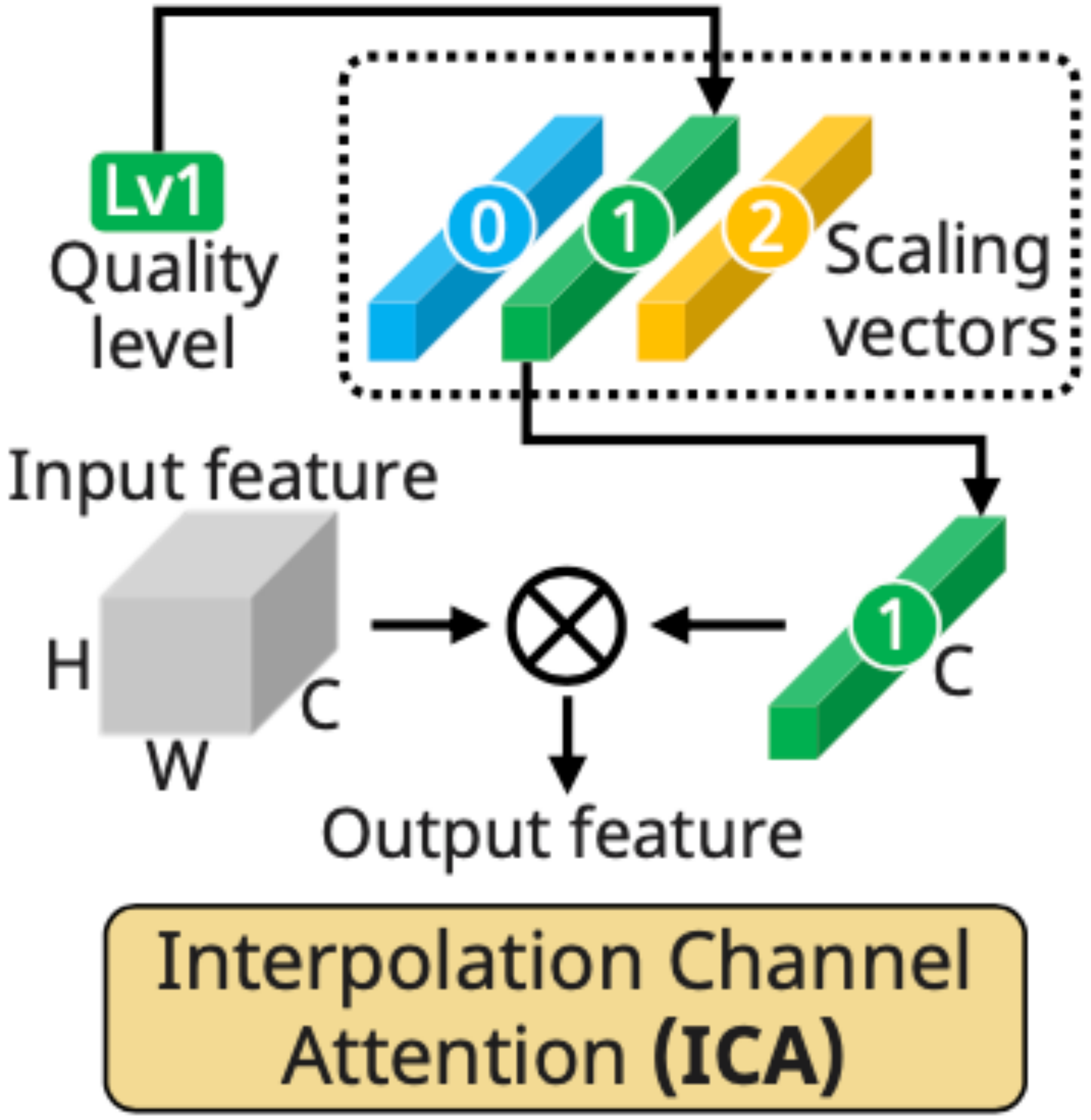
$$\mathcal{L}_{1st} = \lambda_R^{(q)} R(\hat{\boldsymbol{y}}_q) + \lambda_d d(\boldsymbol{x}, \hat{\boldsymbol{x}}_q) + \mathcal{L}_P(\boldsymbol{x}, \hat{\boldsymbol{x}}_q)$$

bit rate, MSE, LPIPS

$$\mathcal{L}_{2nd} = \lambda_R^{(q)} R(\hat{\boldsymbol{y}}_q) + \lambda_d d(\boldsymbol{x}, \hat{\boldsymbol{x}}_q) + \beta(\lambda_P \mathcal{L}_P(\boldsymbol{x}, \hat{\boldsymbol{x}}_q) + \lambda_{\text{adv}} \mathcal{L}^G_{\text{HRRGAN}})$$

bit rate, MSE, LPIPS, adversarial loss

To control the rate:
Insert Interpolation Channel Attention Layers



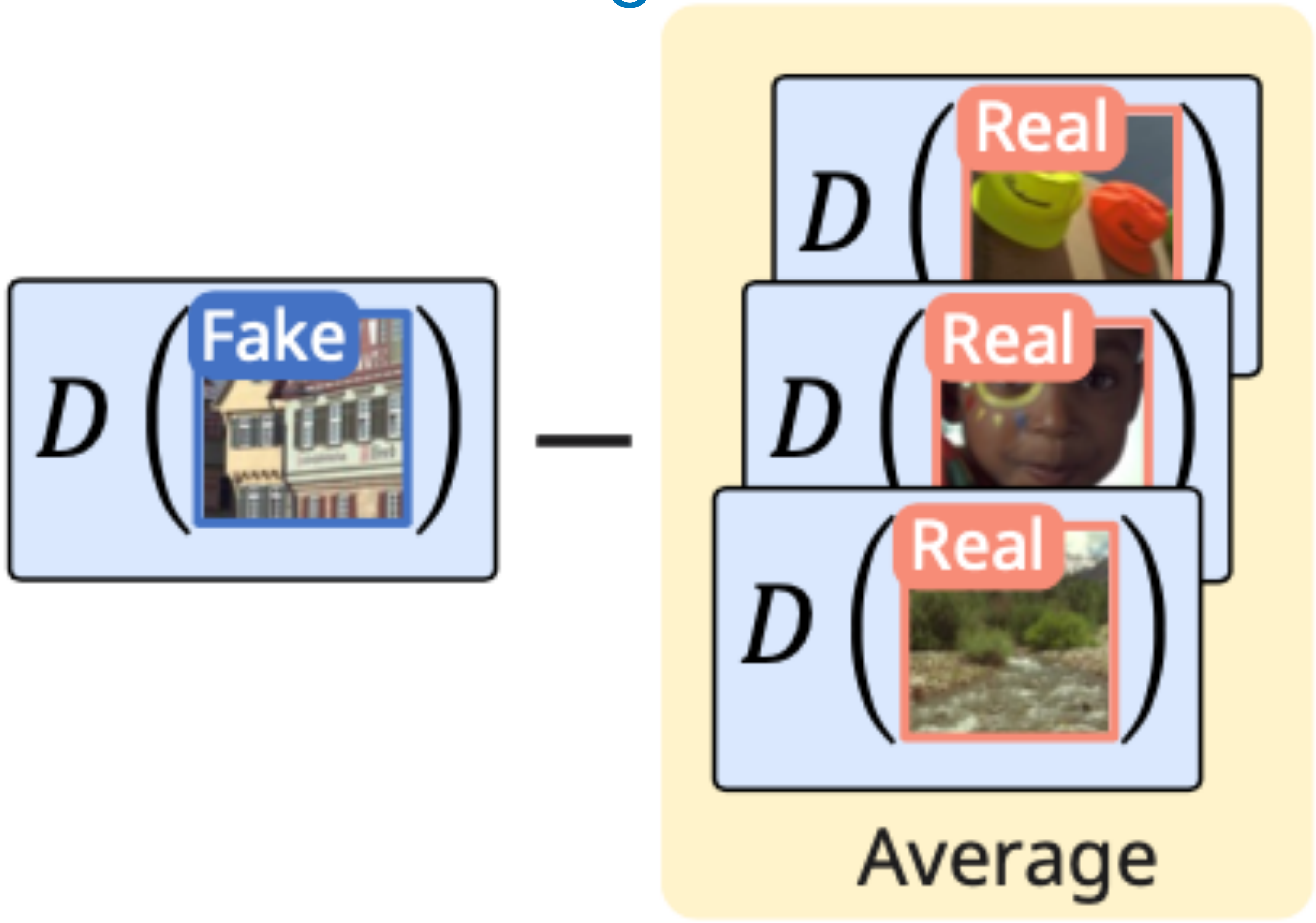
This page till the end: from Iwai et al. [*thisp]

Discriminator - RaGAN

Relativistic Average GAN

$$
\begin{aligned}
p_r(x_r, x_f) &= \sigma(D(x_r) - \mathbb{E}_{x_f}[D(x_f)]) \\
p_f(x_r, x_f) &= \sigma(D(x_f) - \mathbb{E}_{x_r}[D(x_r)]) \\
\mathcal{L}^G_{\text{RaGAN}} &= -\log p_f(x_r, x_f) - \log(1 - p_r(x_r, x_f)) \\
\mathcal{L}^D_{\text{RaGAN}} &= -\log p_r(x_r, x_f) - \log(1 - p_f(x_r, x_f)),
\end{aligned}
$$

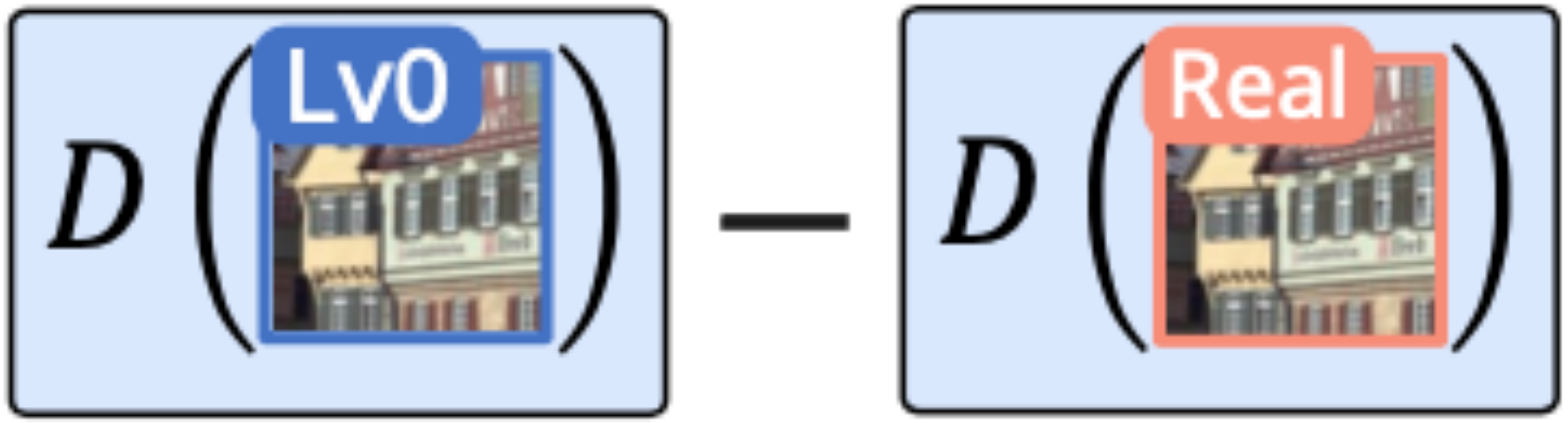
not aligned



Discriminator - RGAN

Relativistic GAN

aligned



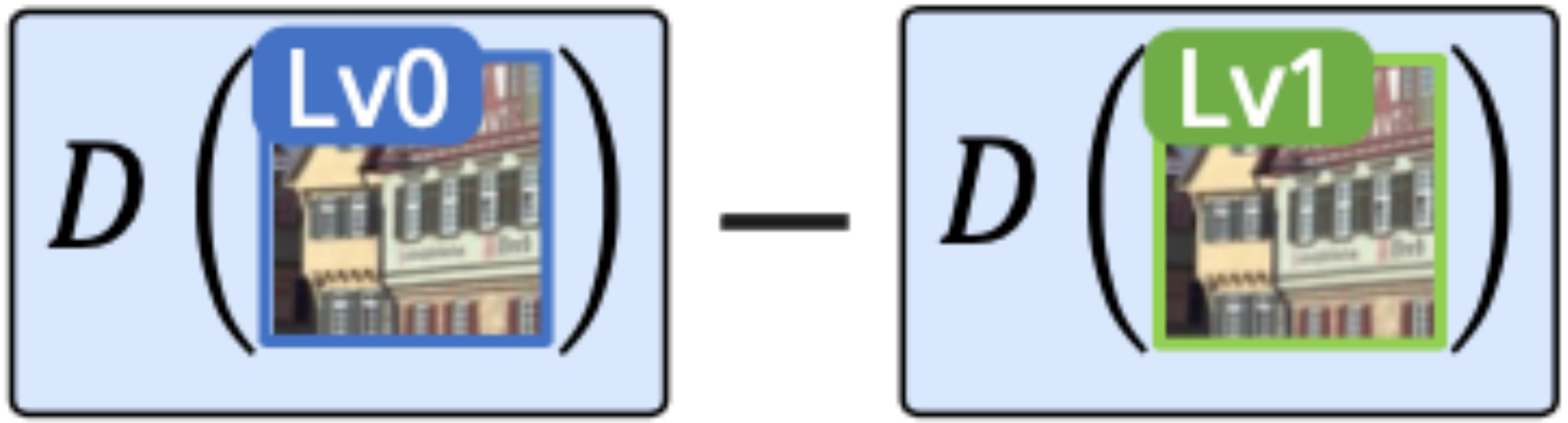
$$\mathcal{L}^{G}_{\text{RGAN}} = -\log \sigma(D(x_f) - D(x_r))$$

$$\mathcal{L}^{D}_{\text{RGAN}} = -\log \sigma(D(x_r) - D(x_f)).$$

Discriminator - HRRGAN

Higher Rate Relativistic GAN

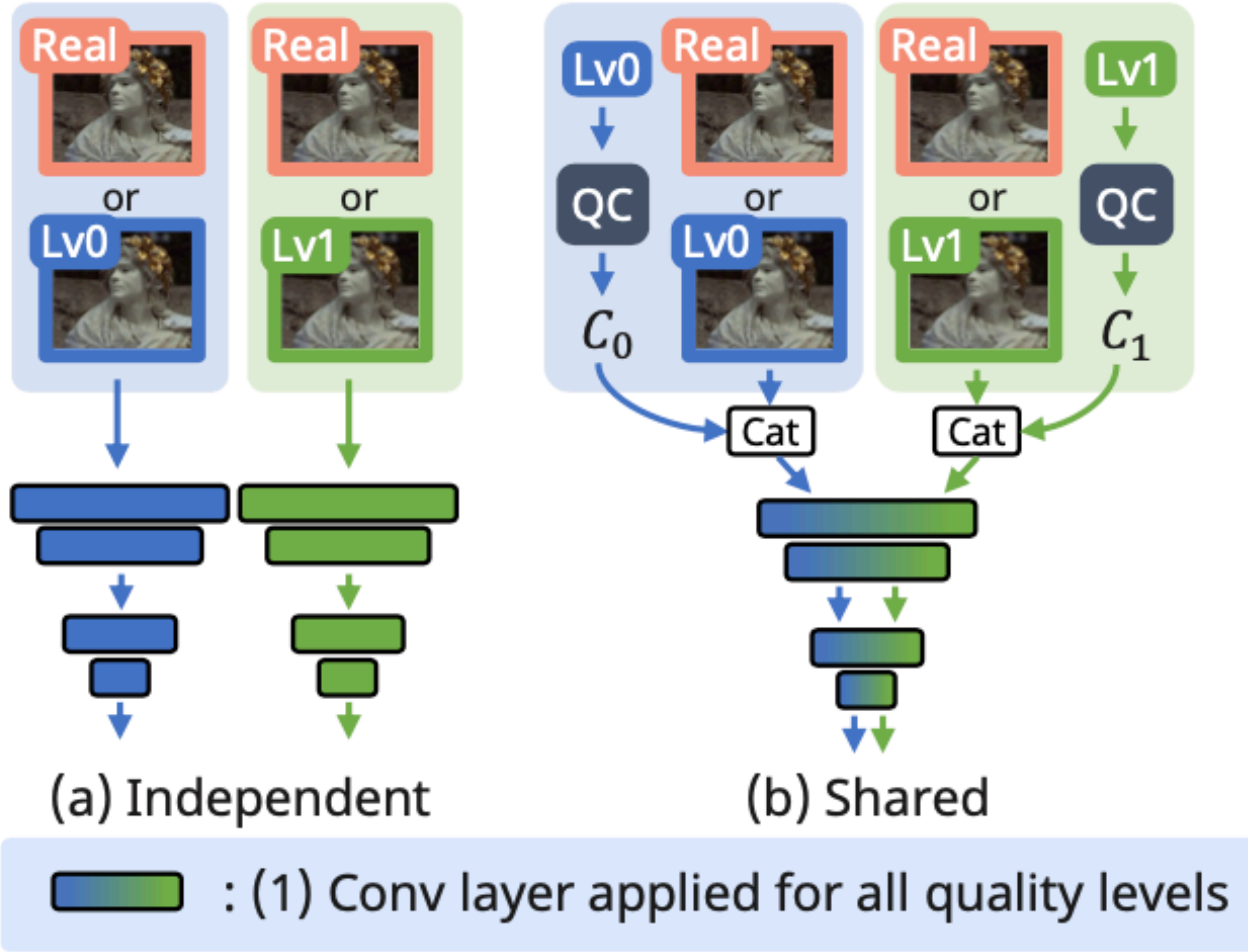
To avoid over-penalty on realism



$$\mathcal{L}^{G}_{\text{HRRGAN}} = -\log\sigma(D(\hat{\boldsymbol{x}}_q) - \text{sg}(D(\hat{\boldsymbol{x}}_{q+1})))$$
$$\mathcal{L}^{D}_{\text{HRRGAN}} = -\log\sigma(D(\boldsymbol{x}) - D(\hat{\boldsymbol{x}}_q)),$$

sg: stop gradient operation

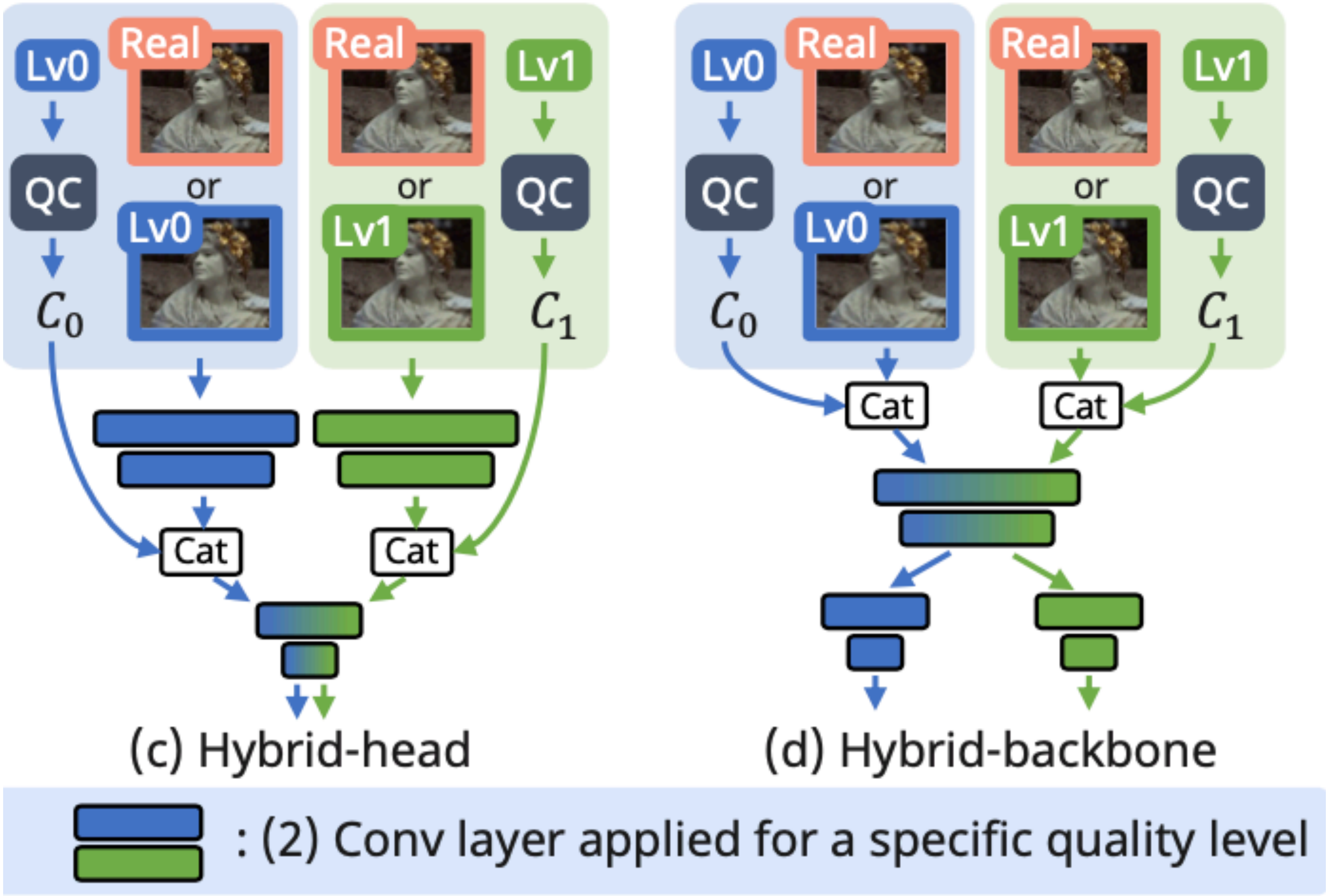
Independent vs Shared Discriminator



Hybrid Discriminator

backbone: extract and encode features

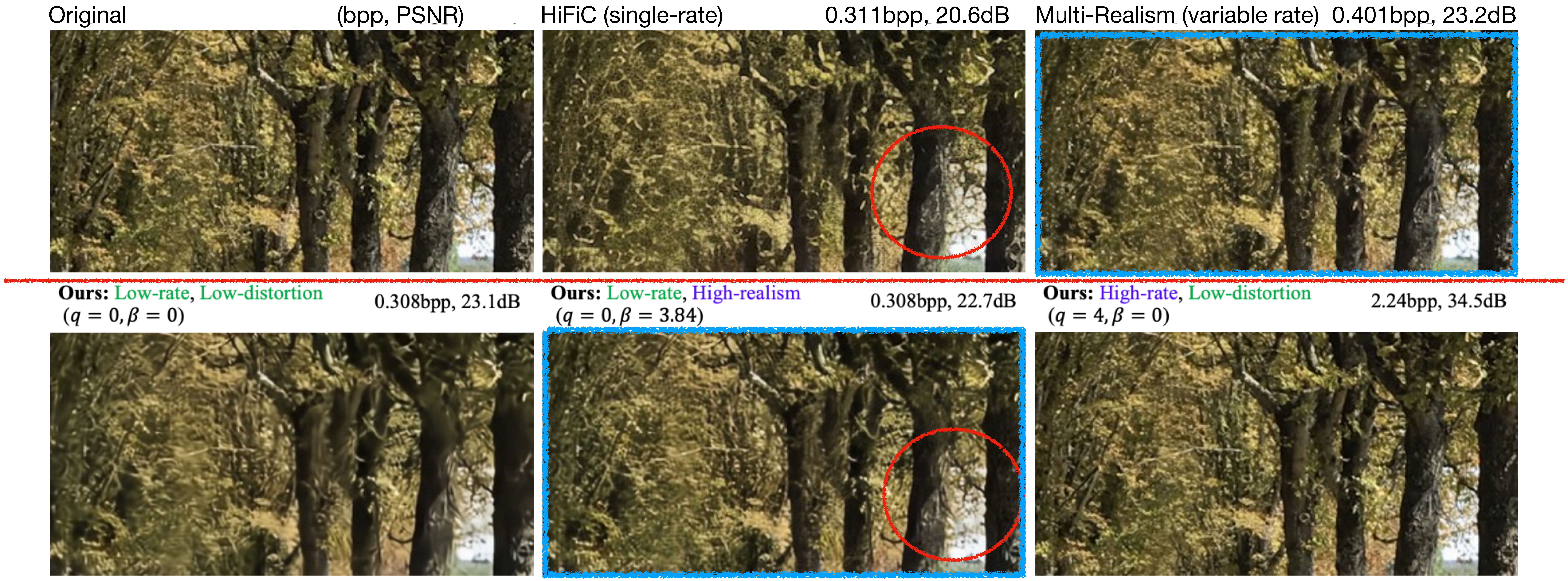
head: produce prediction



(c) Hybrid-head

(d) Hybrid-backbone

Experimental Results: Compare with Generative Models



Original (bpp, PSNR)

HiFiC (single-rate) 0.311bpp, 20.6dB

Multi-Realism (variable rate) 0.401bpp, 23.2dB

Ours: Low-rate, Low-distortion ($q = 0, \beta = 0$) 0.308bpp, 23.1dB

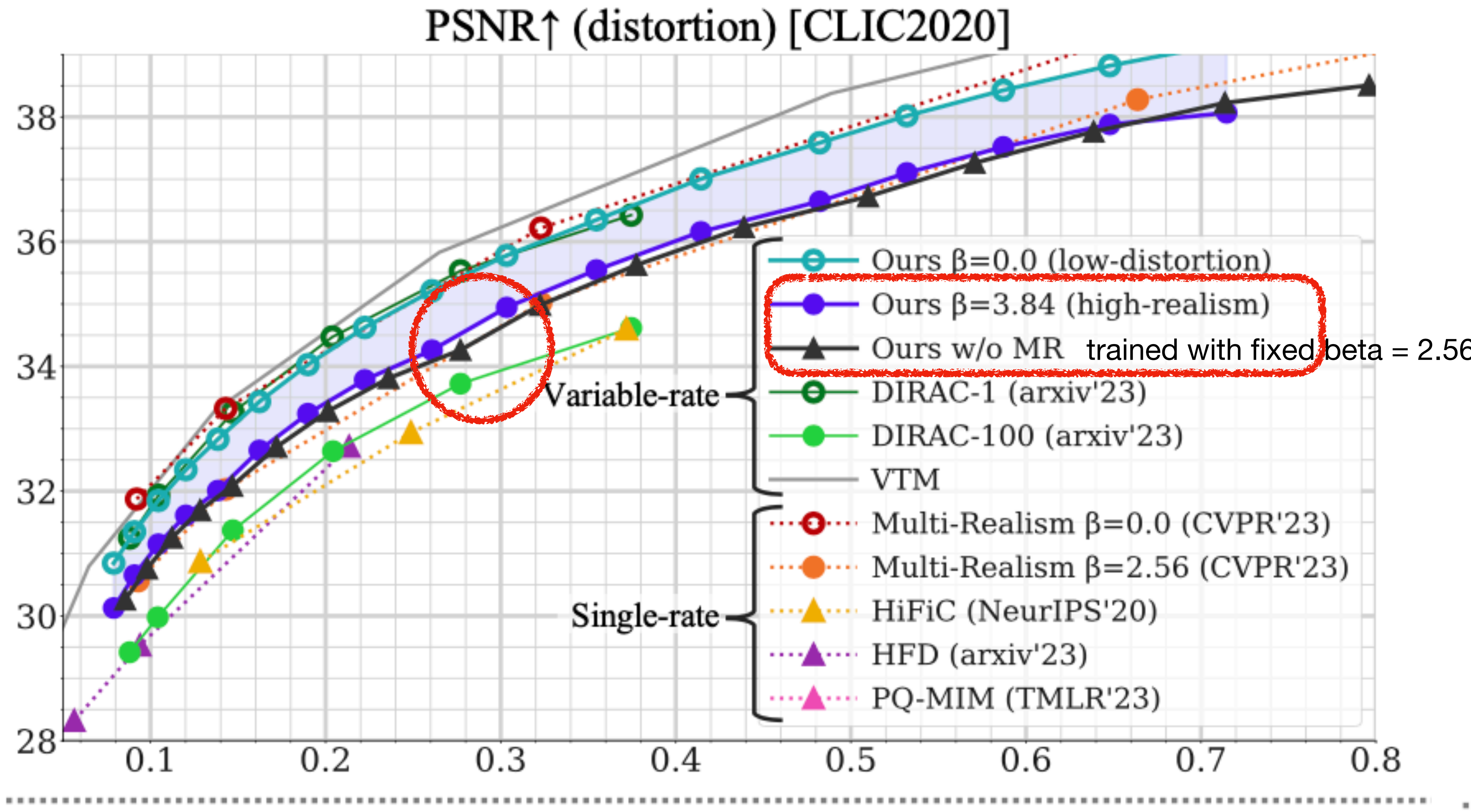
Ours: Low-rate, High-realism ($q = 0, \beta = 3.84$) 0.308bpp, 22.7dB

Ours: High-rate, Low-distortion ($q = 4, \beta = 0$) 2.24bpp, 34.5dB

better texture!

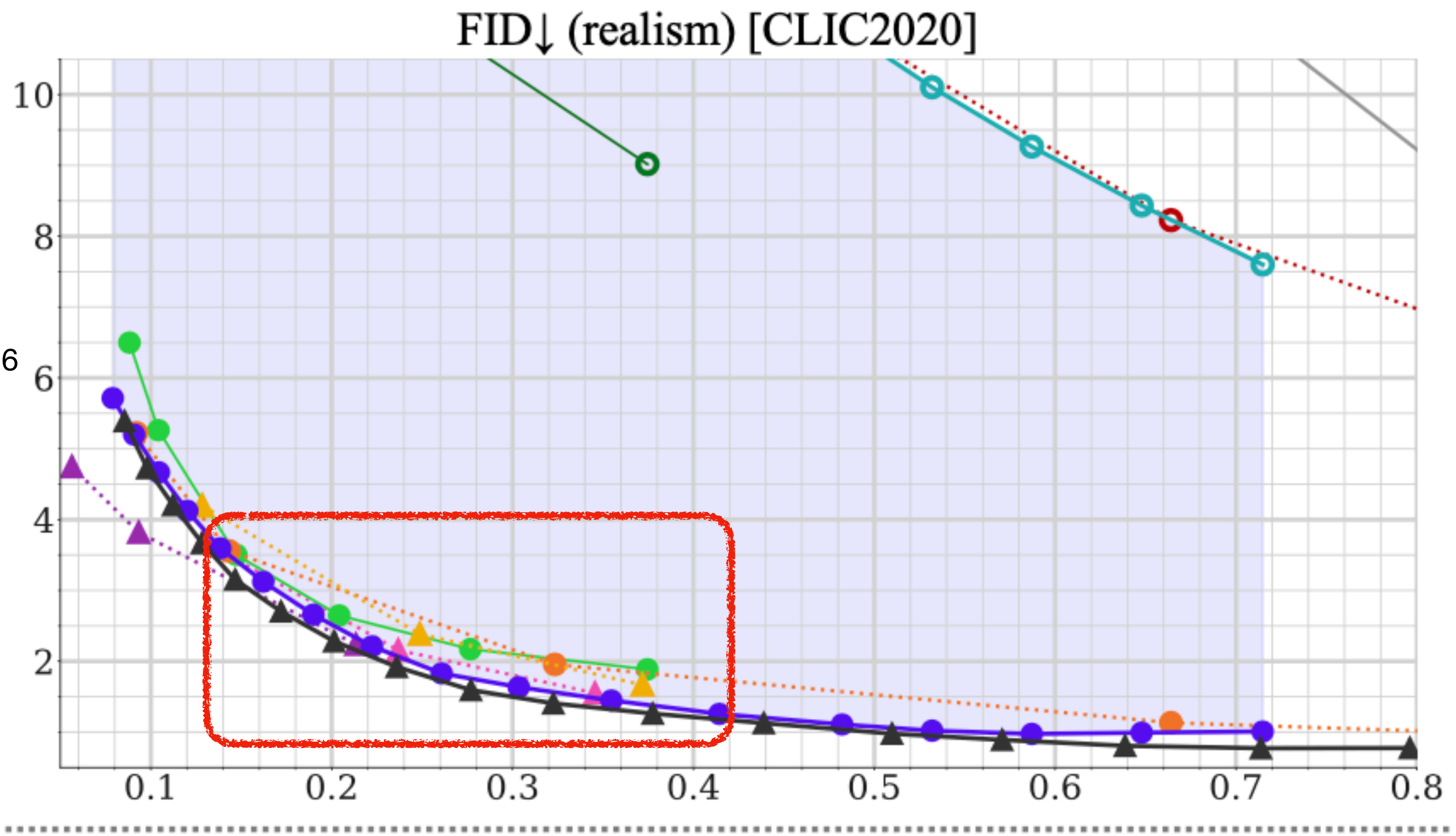
Comparable realism with lower bit rate!

Quantitative Evaluation



trained with fixed beta = 2.56

bpp(bit rate)

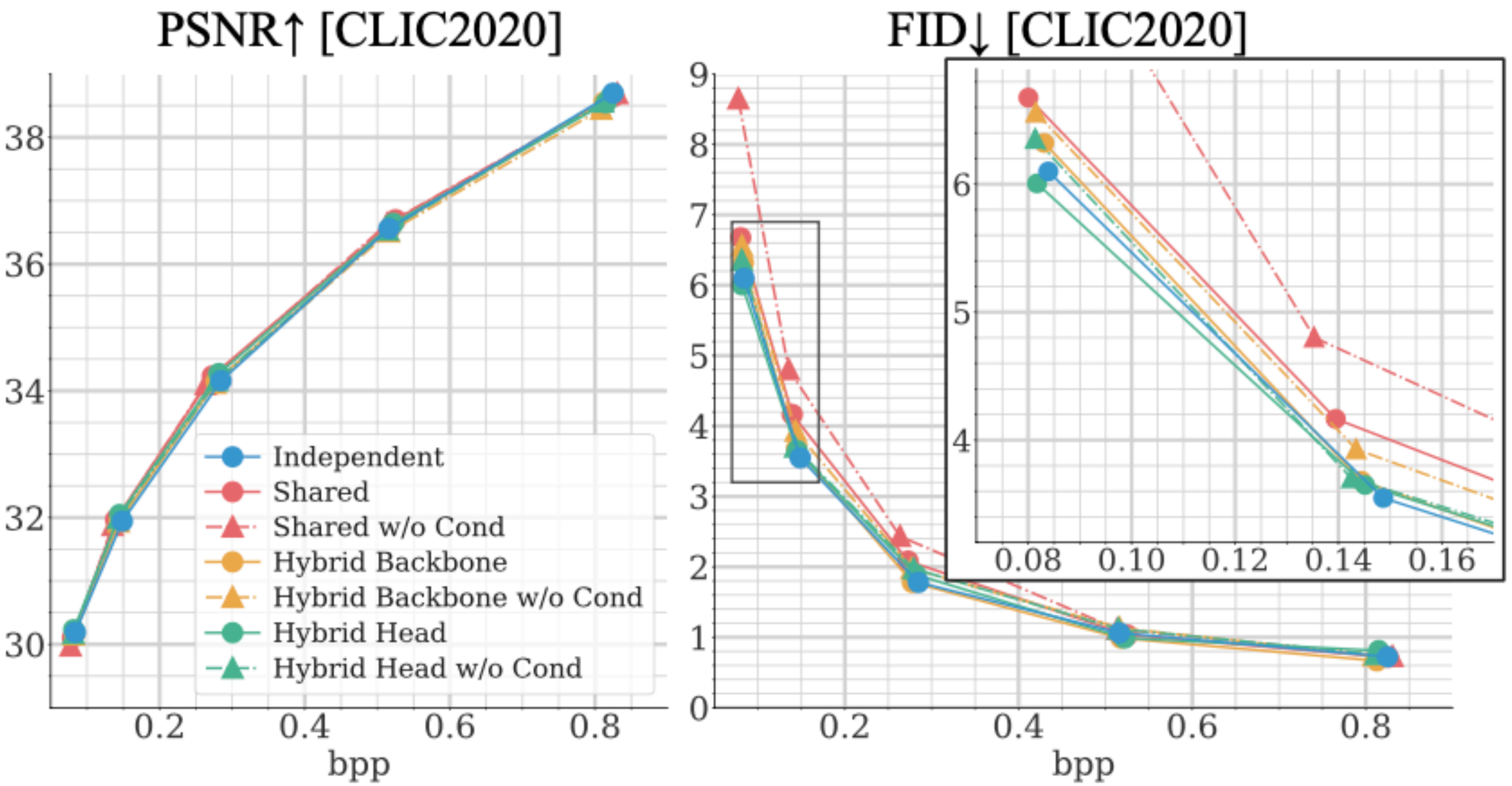


bpp(bit rate)

Quantitative Evaluation

- Perform fine rate-tuning
- high realism model: surpassed DIRAC on both indicators

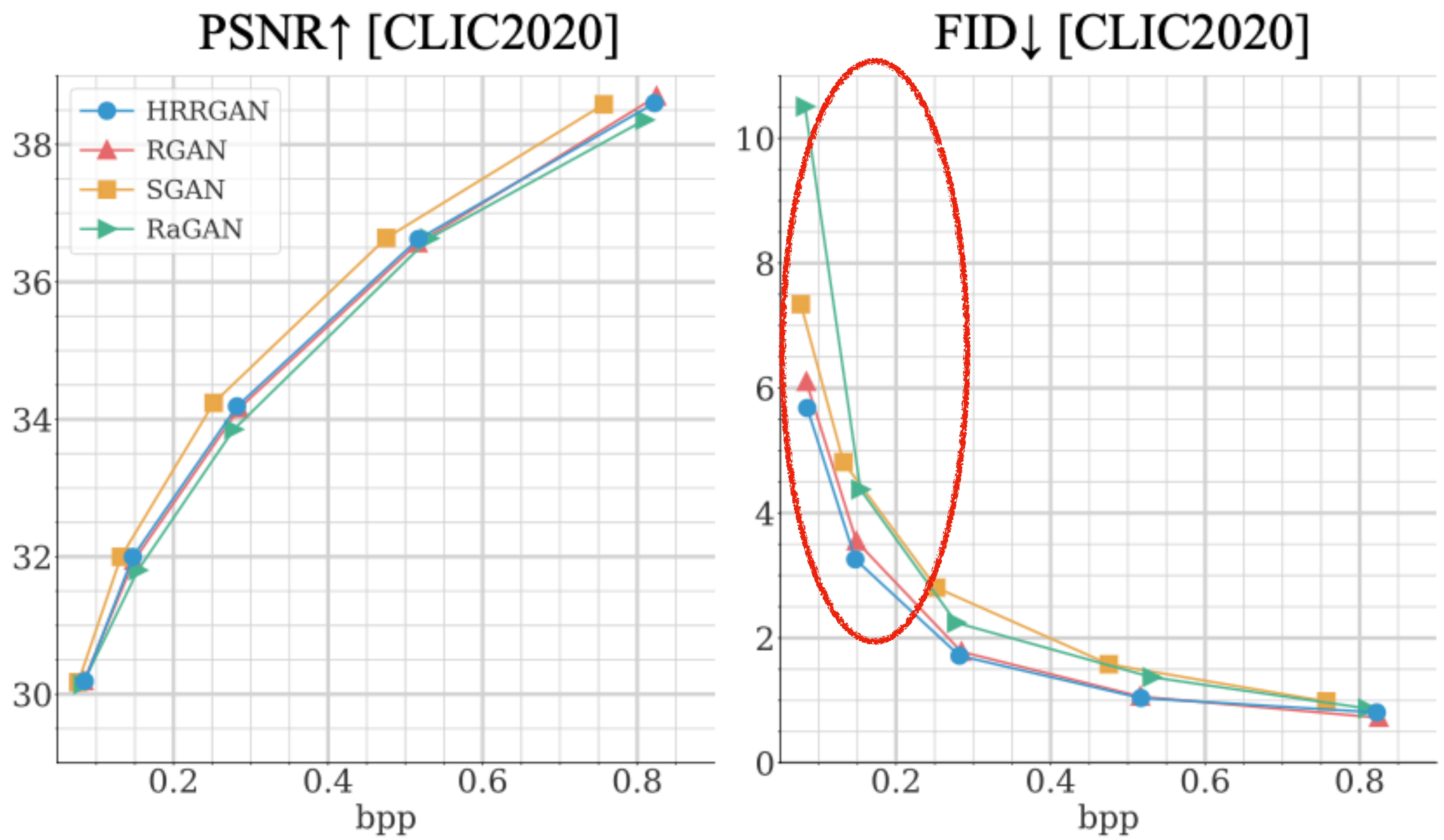
Results of Different Discriminator designs



(a) Results on $Q = 5$

- Hybrid discriminators outperformed shared discriminators in FID
- Quality-level specific layers are beneficial

Effect of HRRGAN



- Average calculation harms performance

Trained with fixed beta = 2.56

RGAN: Relativistic GAN

SGAN: Standard GAN

RaGAN: Relativistic Average GAN

HRRGAN: Higher Rate Relativistic GAN

Limitation

- Control the rate and realism uniformly

  cannot perform precise (e.g. pixel-level) control

References

[*multi-real] Agustsson, Eirikur, et al. "Multi-realism image compression with a conditional generator." *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 2023.

[*hyperprior] Ballé, Johannes, et al. "Variational image compression with a scale hyperprior." *arXiv preprint arXiv:1802.01436* (2018).

[*measure] Becker Axel. "A review on image distortion measures." (2000).

[*realism] Fan, Shaojing, et al. "Image visual realism: From human perception to machine computation." *IEEE transactions on pattern analysis and machine intelligence* 40.9 (2017): 2180-2193.

[*thisp] Iwai Shoma, Tomo Miyazaki, and Shinichiro Omachi. "Controlling Rate, Distortion, and Realism: Towards a Single Comprehensive Neural Image Compression Model." *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*. 2024.

[*diffPSNR] Loukil Habiba, Moez Hadj Kacem, and Mohamed Salim Bouhlel. "A new image quality metric using system visual human characteristics." *International Journal of Computer Applications* 60.6 (2012).

[*sameMSE] Wang, Zhou, et al. "Image quality assessment: from error visibility to structural similarity." *IEEE transactions on image processing* 13.4 (2004): 600-612.

[*tradeoff] Yochai Blau and Tomer Michaeli. Rethinking lossy compression: The rate-distortion-perception tradeoff. In *Proceedings* of the 36th International Conference on Machine Learning*(ICML)*, 2019.