

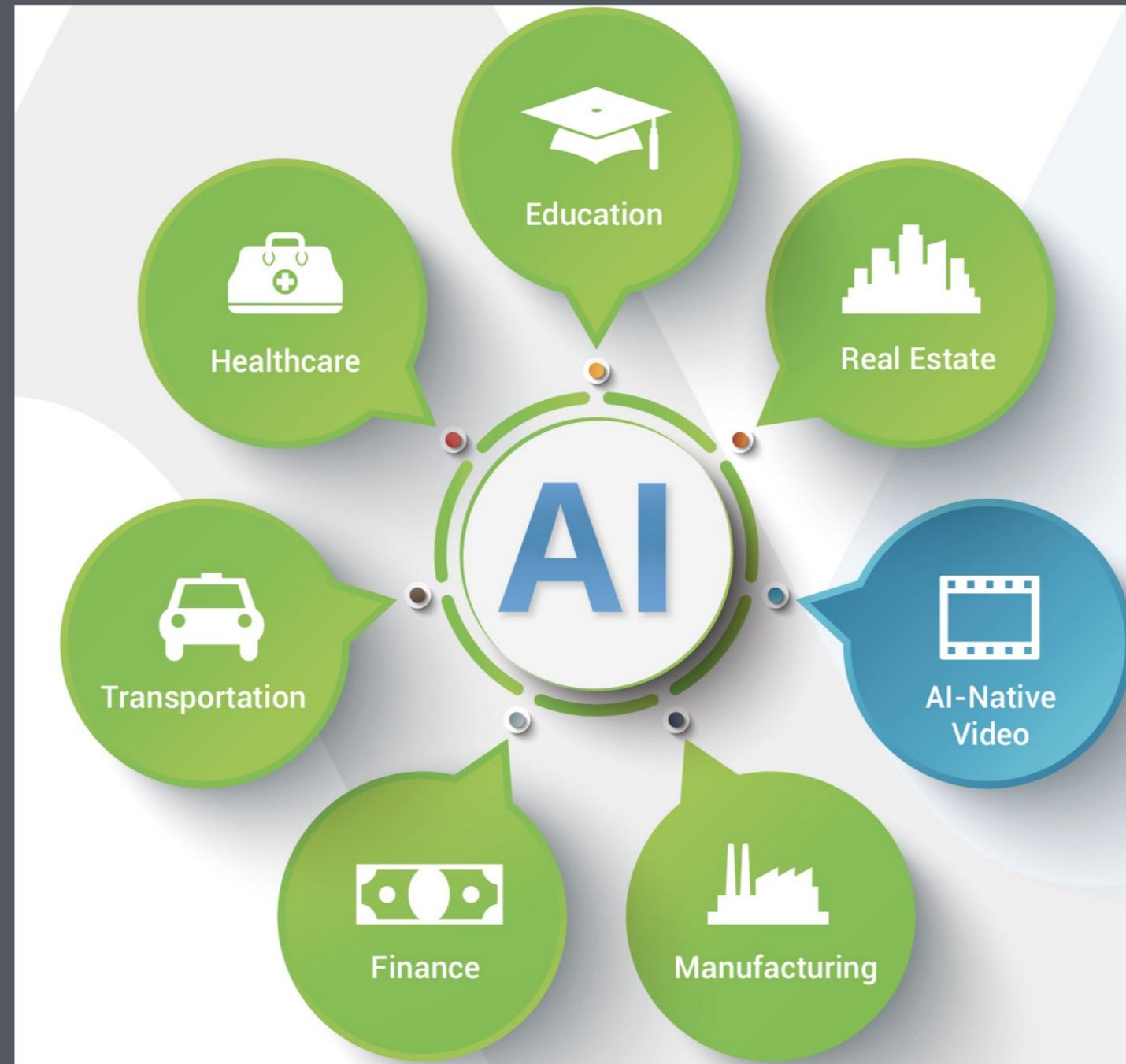
Reinventing video compression with Deep Neural Networks

Opportunities and Challenges

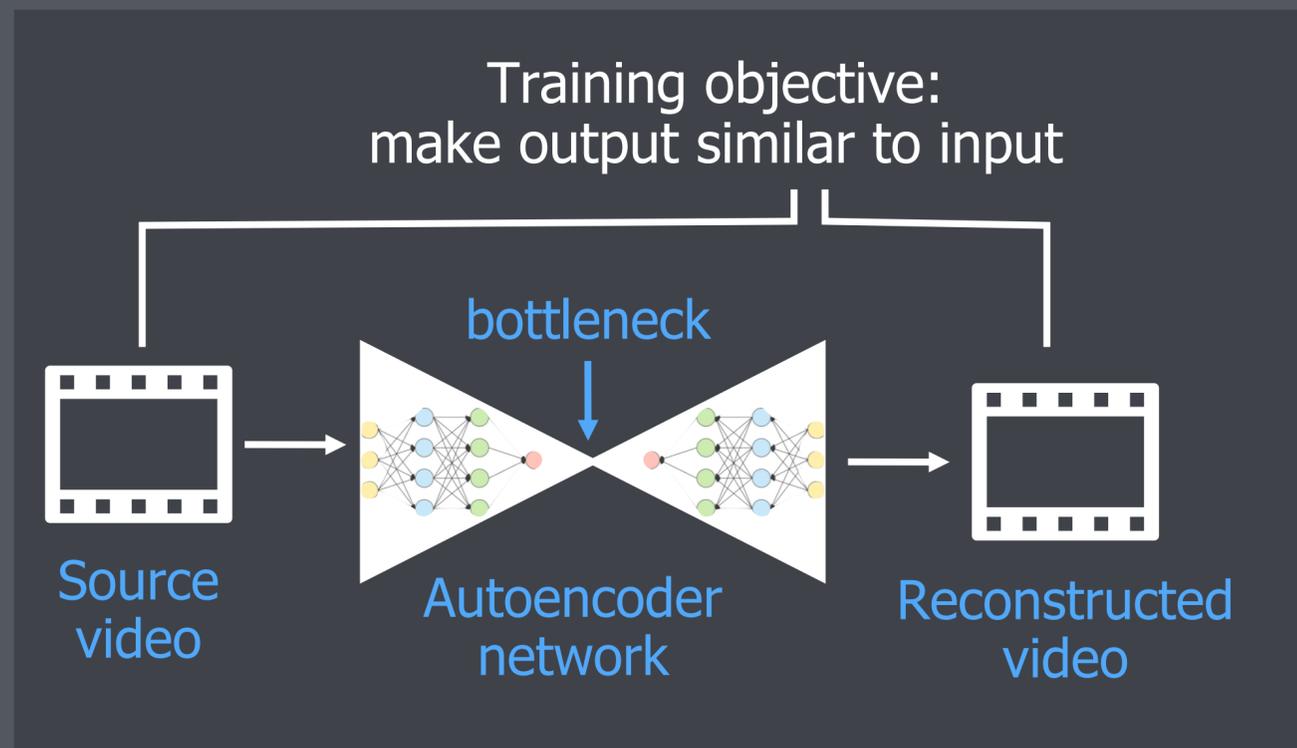
Lubomir Bourdev
Co-founder / CEO



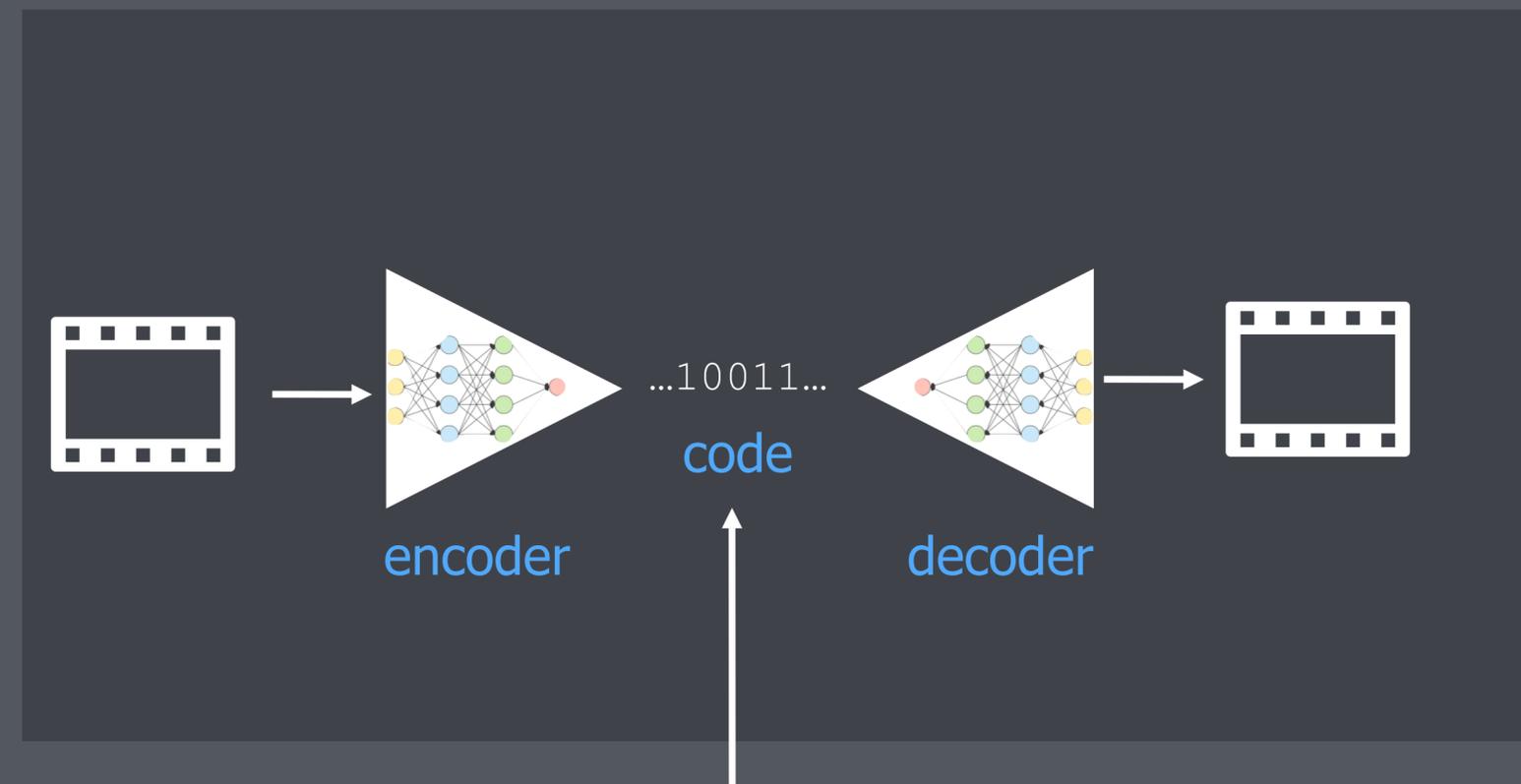
AI Has made tremendous progress in the past decade



Training setup



Deploy setup



AI-Native training automatically tunes millions of parameters

Traditional codecs use a *hard-coded representation* (i.e. H.264).
In AI-Native the representation is *derived during training*

Why Neural Networks are great for compression?

"Nearby pixels are similar"

"Changes are often due to motion"

"Nearby pixels tend to move together"

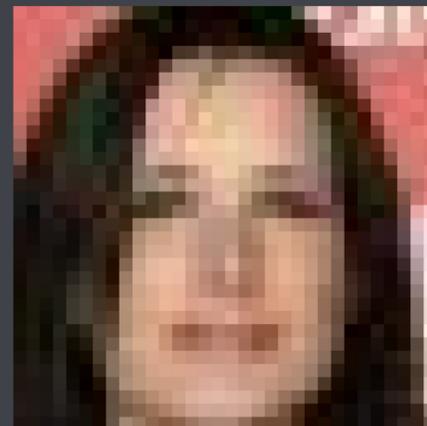


Traditional codecs only capture high level principles

But the natural world is very complicated

Neural Networks can "fill in the blanks" much better

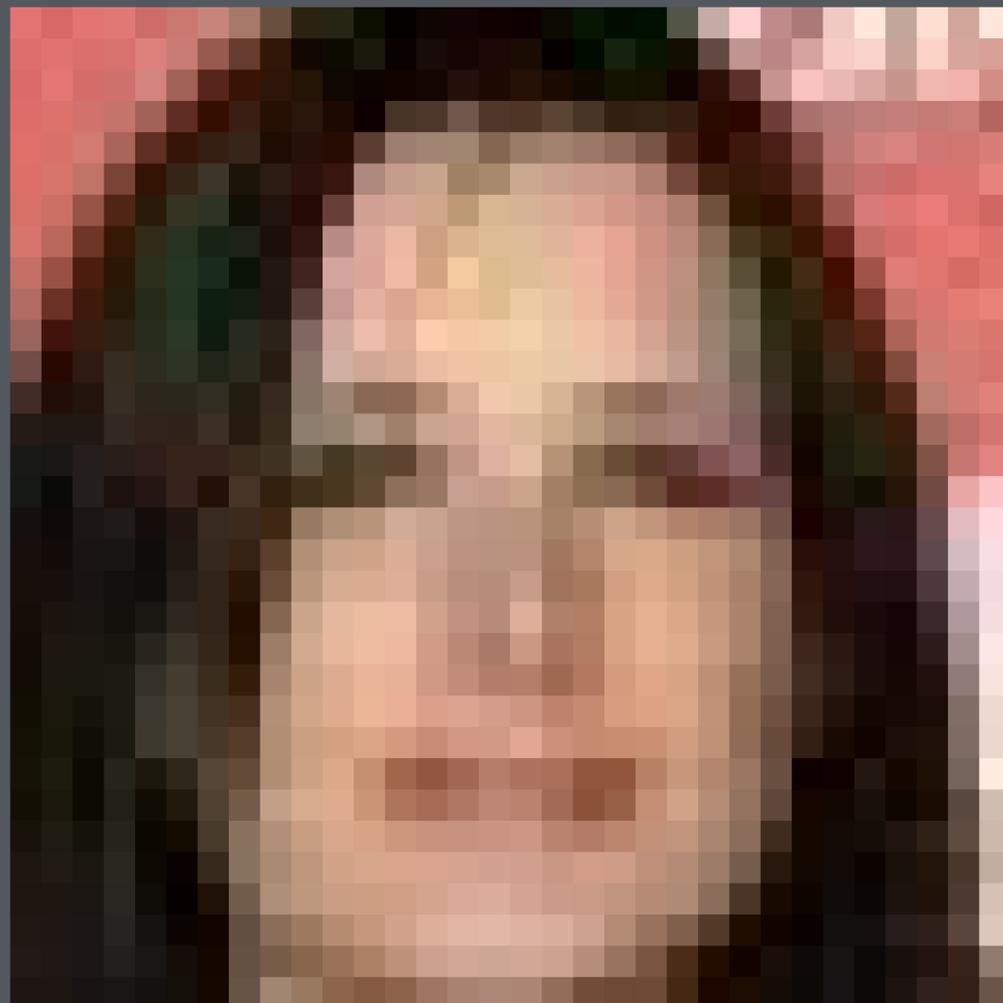
32x32
input



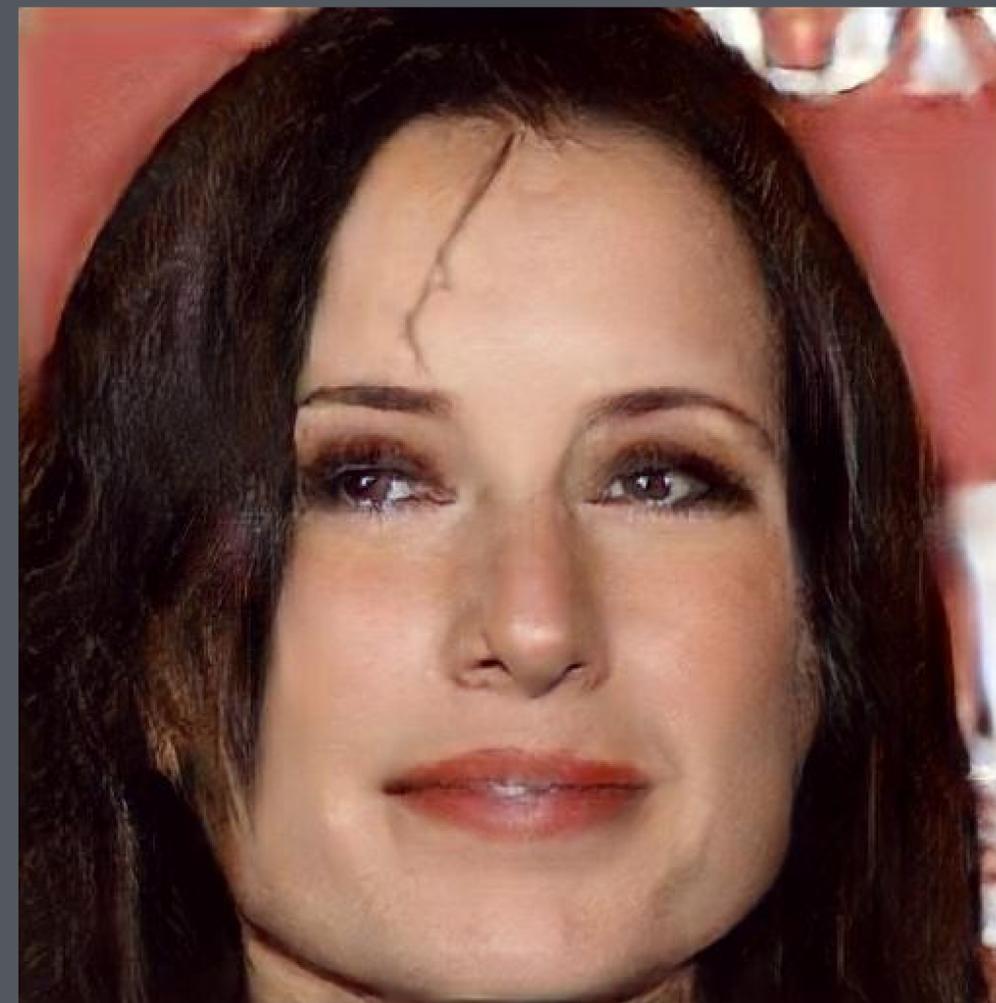
512x512
output

WaveOne superresolution (trained on faces)

WaveOne Super-resolution (Trained on Faces)



32x32 input



512x512 output

Motion represented as quad tree?



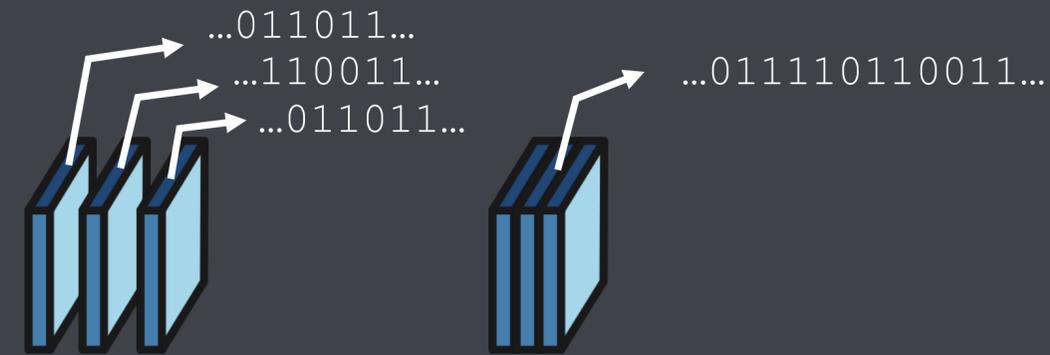
H.265



WaveOne



Frame-by-frame compression?



Single motion field?



And many other ideas not possible with standards

- ▶ Hierarchical representations
- ▶ Associating motion with uncertainty
- ▶ Combining compression and super-resolution
- ▶ Joint compression of motion and residual
- ▶ Explicit handling of occlusion and disocclusion
- ▶ ...

A library of custom-tailored codecs

- ▶ Custom-tailored to domain

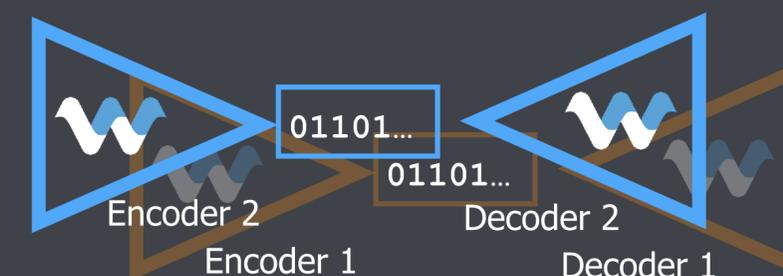


- ▶ Custom-tailored for ultra low latency
Trained to handle missing network packets

- ▶ Custom-tailored for ultra-low bandwidth



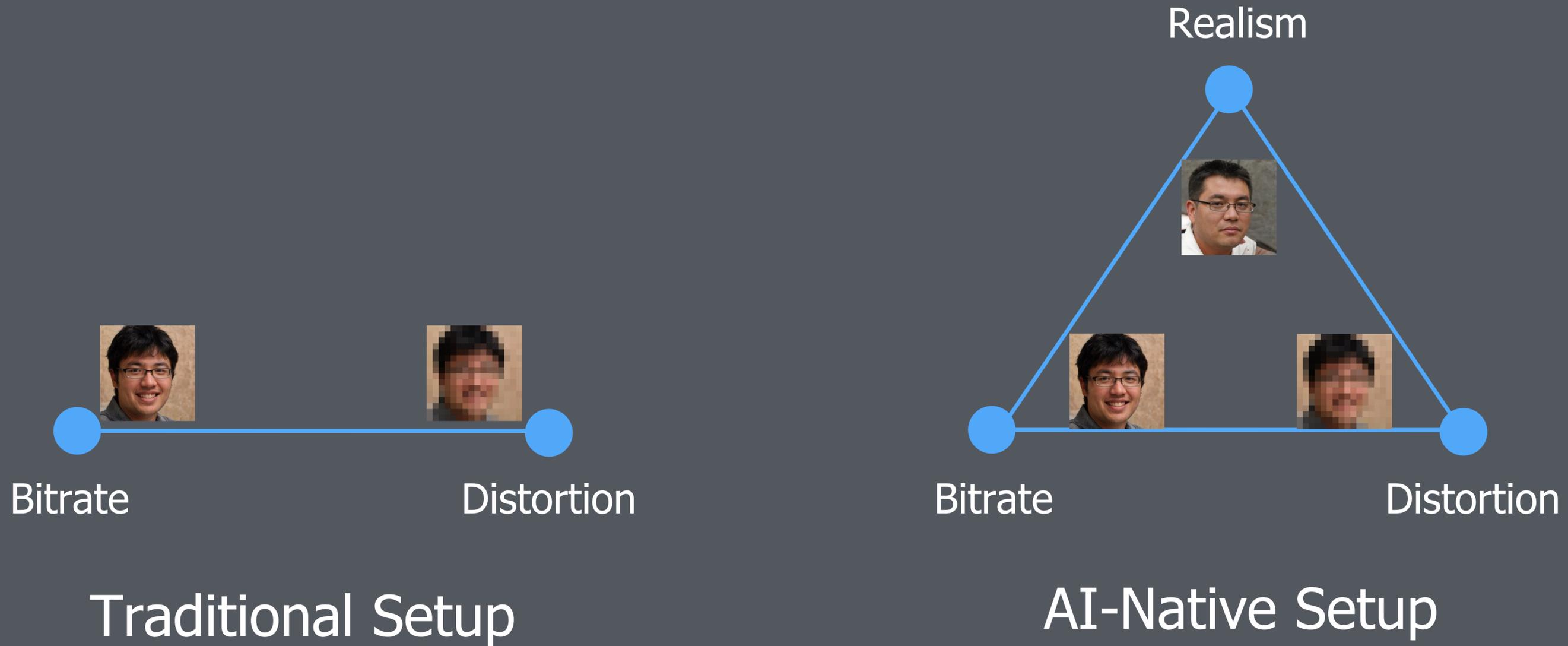
Dynamically update codecs



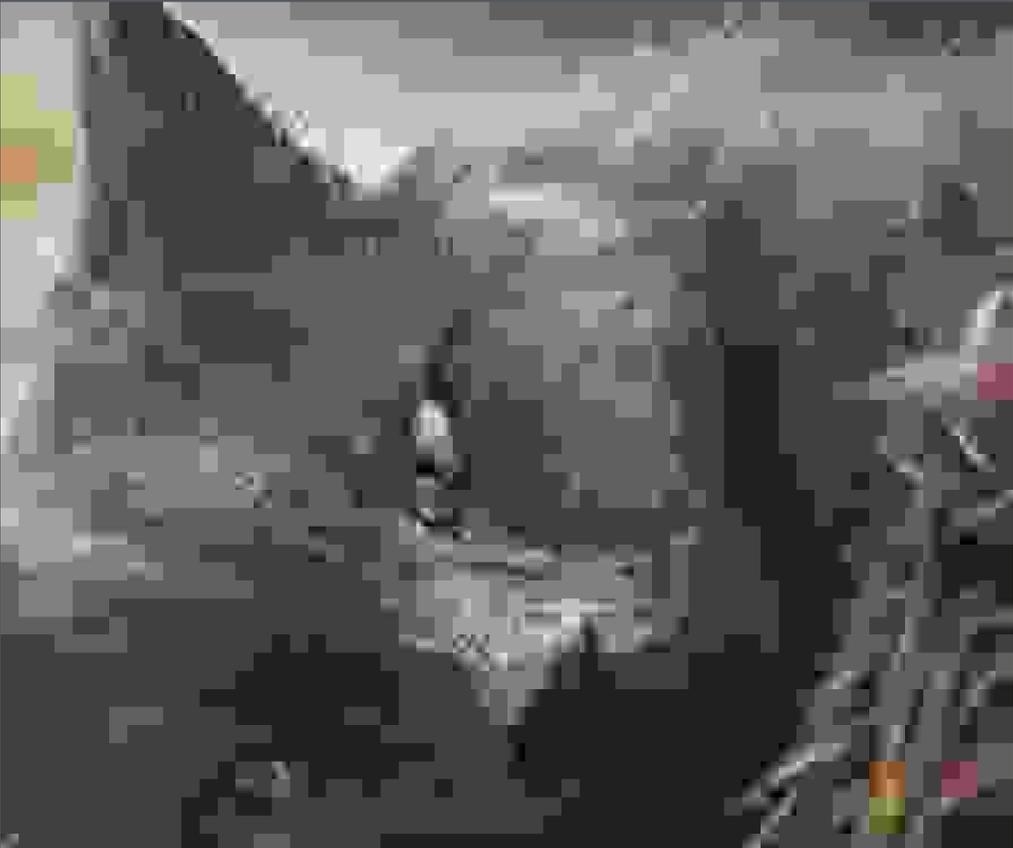
Leverage industry-wide DNN accelerators



Rethinking Lossy Compression objectives



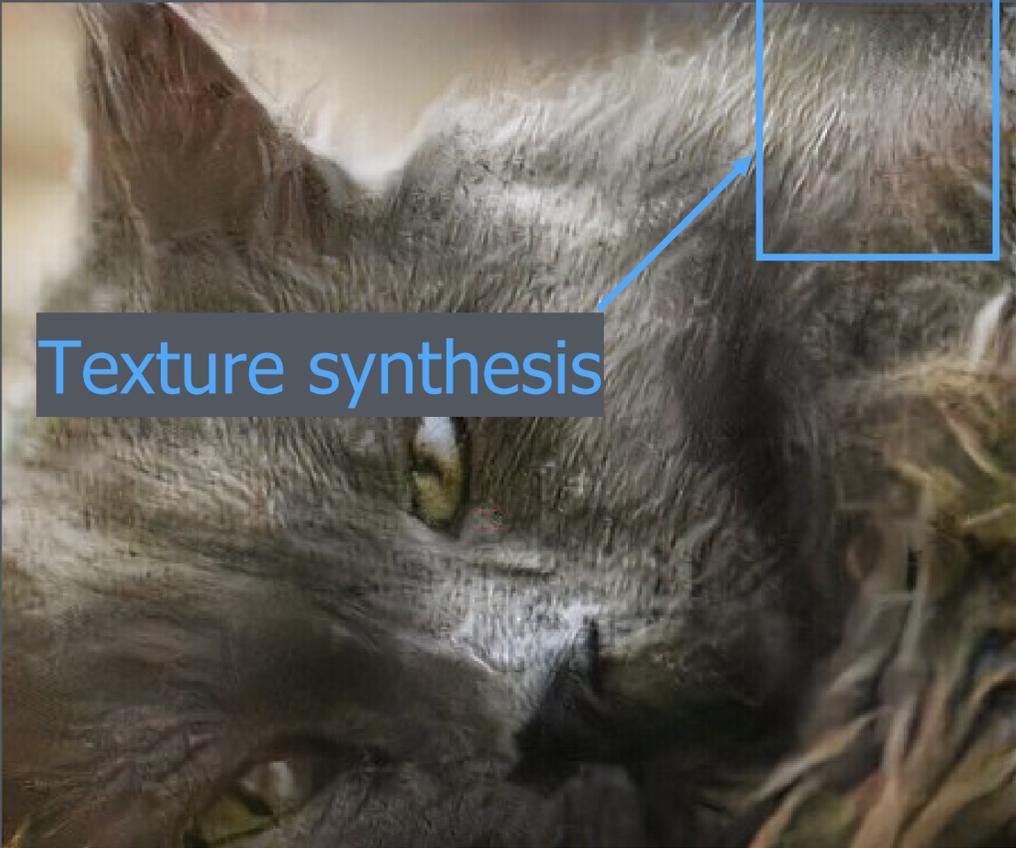
I-frame, 0.098 bits/pixel



JPEG



BPG



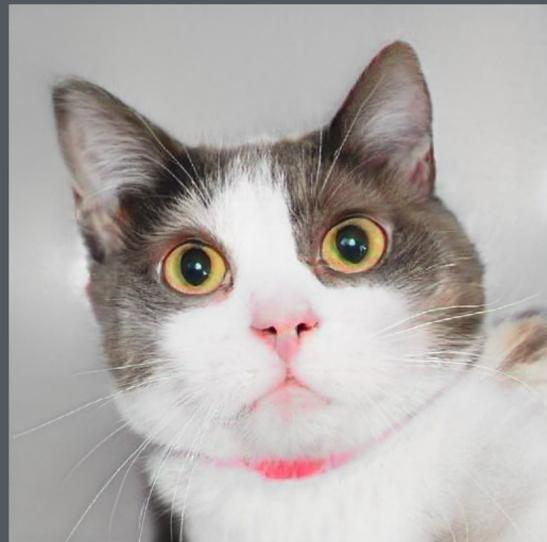
Texture synthesis

WaveOne 2019

Uncanny power of Image synthesis with Neural Networks WaveOne



thispersondoesnotexist.com



thiscatdoesnotexist.com



thishorsedoesnotexist.com

Rethinking Perceptual Quality Metrics

- ▶ Existing metrics (PSNR, MS-SSIM, VMAF) are poor proxy for quality

Important



Not important (yet costly)

- ▶ Good metric needs to consider:
 - Semantics
 - Object saliency
 - Attention model
 - Capabilities of human visual system

- ▶ We have a collaboration with academia to build the next generation quality metric



- ▶ Better metric can make AI-Native an order of magnitude better

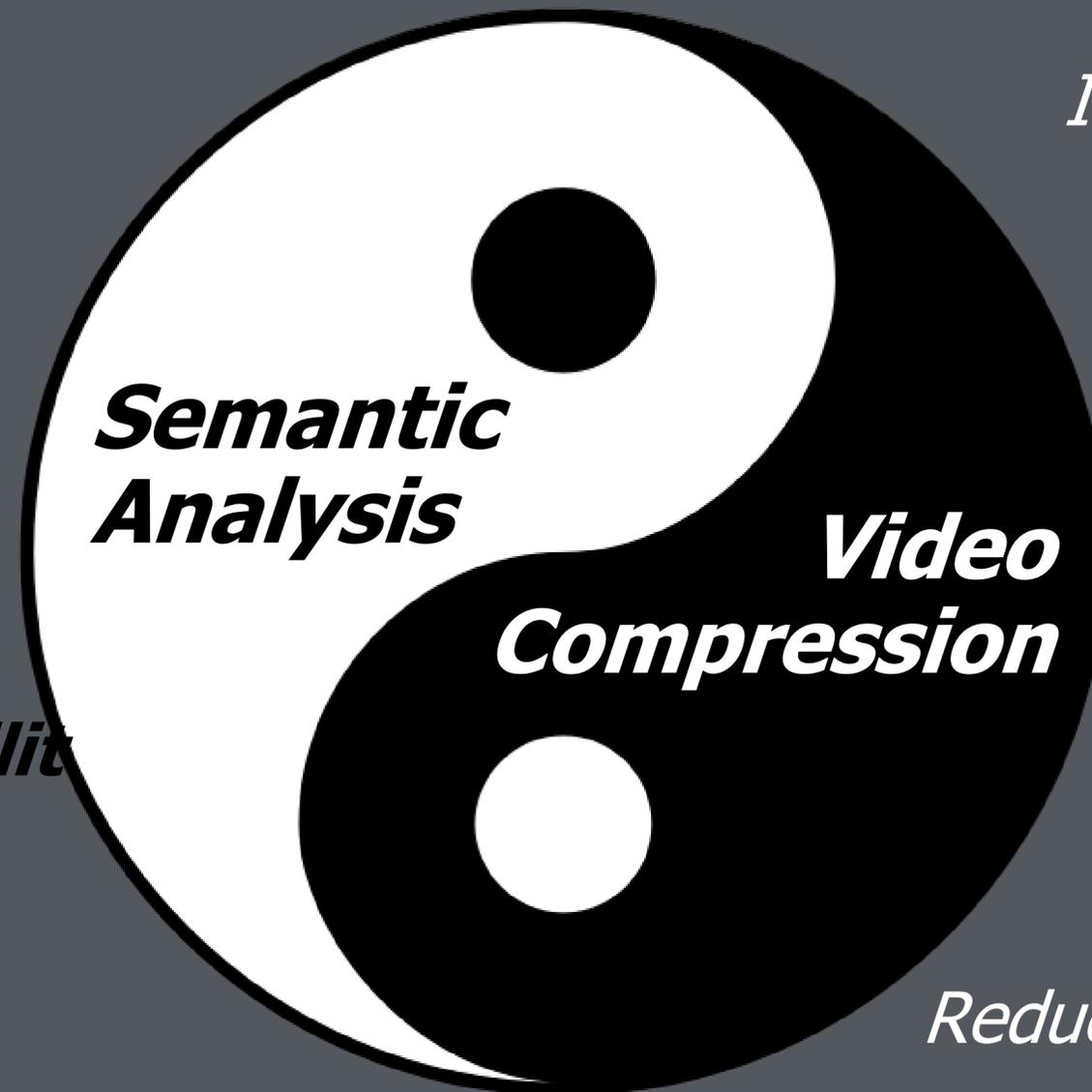
Content ranking

Moderation

Ad revenue

Discoverability

Privacy compliance



Increasing video reach

Cheap archiving

Ultra low latency

Higher video quality

Reducing transmission cost

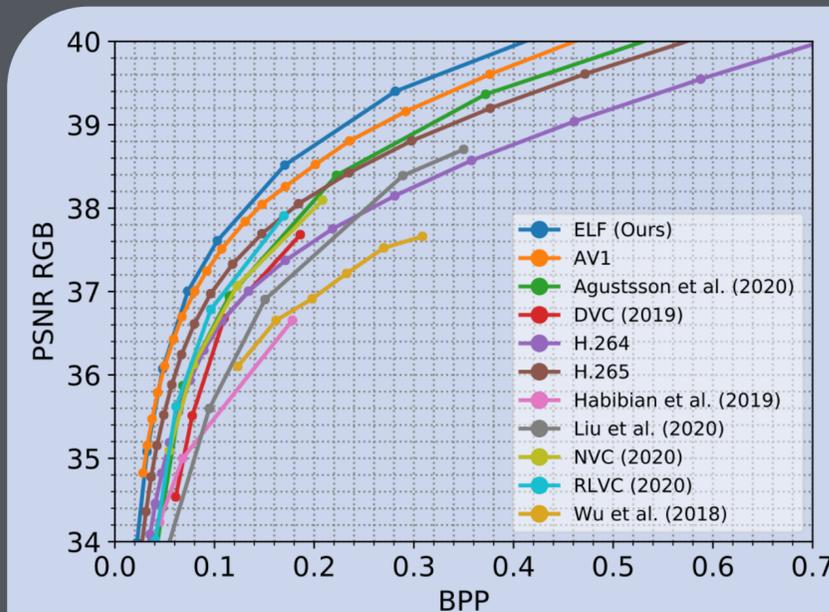
Semantic analysis is built into AI-Native compression

Questions about AI-Native compression

- ▶ Is it competitive vs standards?
- ▶ Is it slow and inefficient?
- ▶ Is it not standards-based?
- ▶ Is there hardware support?
- ▶ Are large companies investing in it?
- ▶ Is it difficult to transition to it?

Questions about AI-Native compression

- ▶ **Is it competitive vs standards?**
- ▶ Is it slow and inefficient?
- ▶ Is it not standards-based?
- ▶ Is there hardware support?
- ▶ Are large companies investing in it?
- ▶ Is it difficult to transition to it?

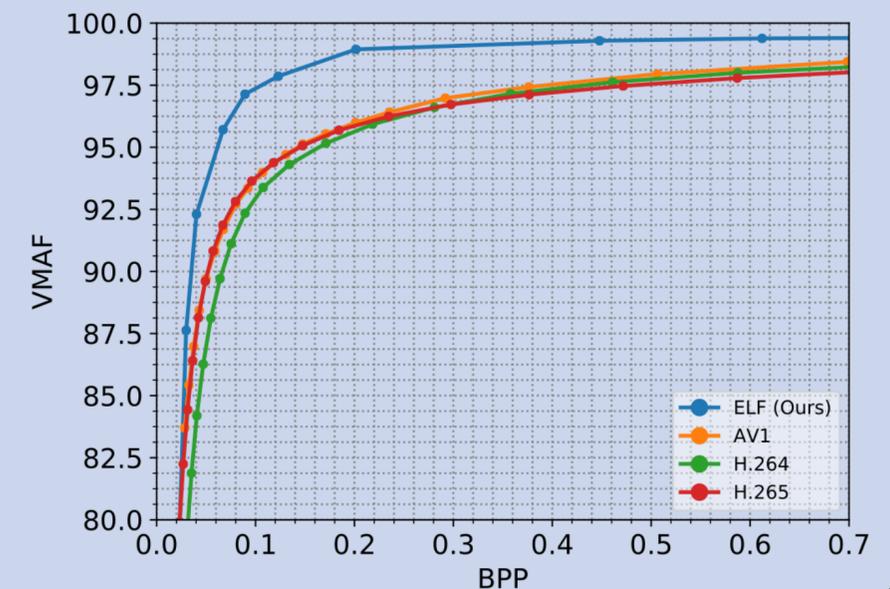
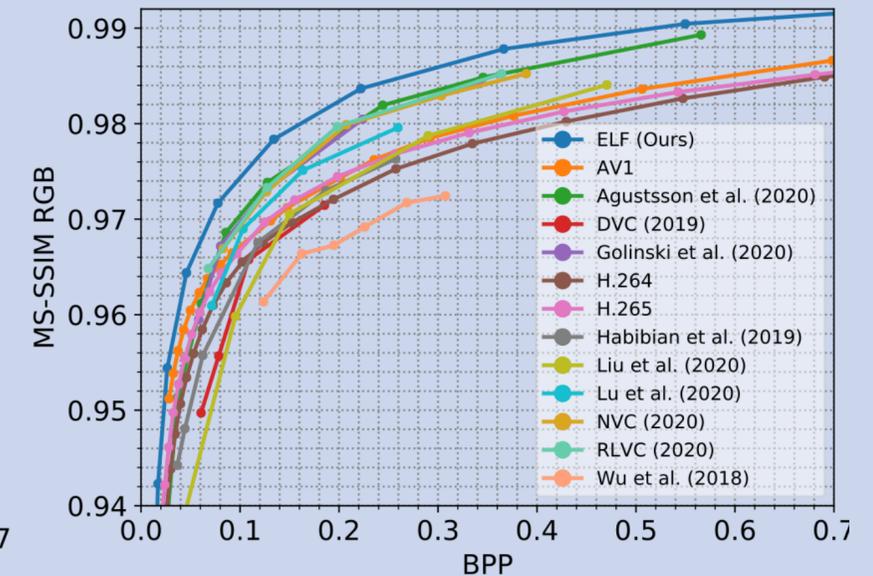


[Rippel et al, ICCV 2021]

Dataset: UVG

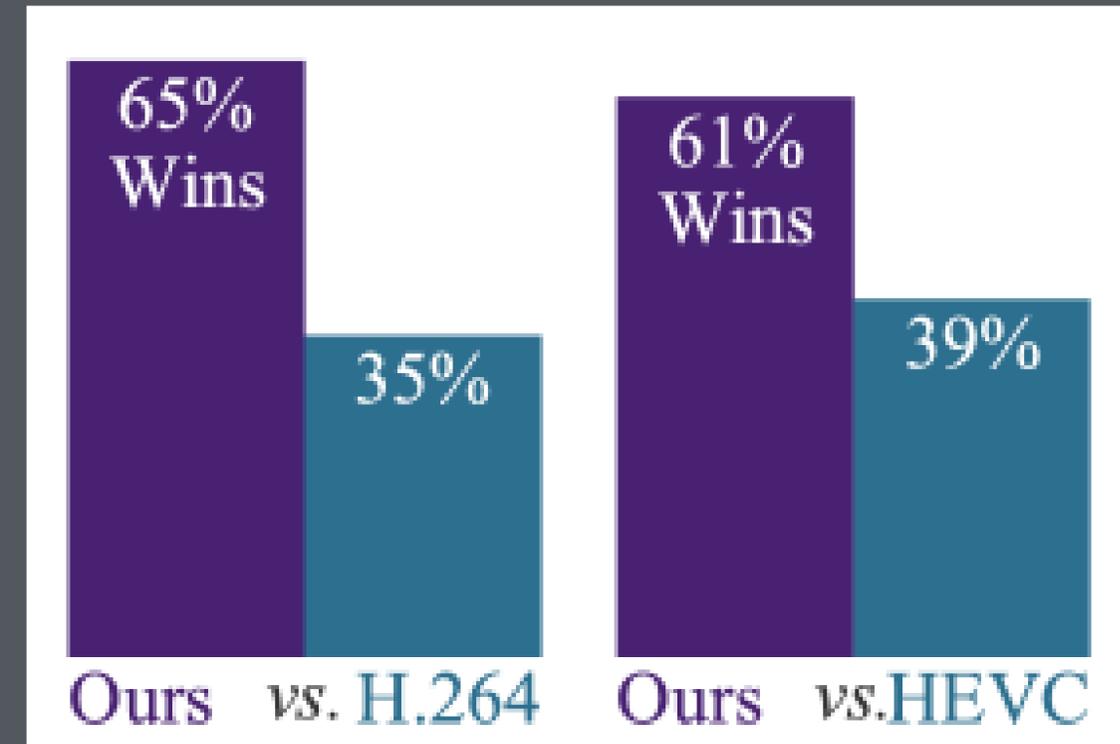
Encoding: Low-latency

See wave.one/elf



Questions about AI-Native compression

- ▶ **Is it competitive vs standards?**
- ▶ Is it slow and inefficient?
- ▶ Is it not standards-based?
- ▶ Is there hardware support?
- ▶ Are large companies investing in it?
- ▶ Is it difficult to transition to it?



User study results on MCL-JCV dataset
[Mentzer et al, arXiv 2021]

Questions about AI-Native compression

- ▶ Is it competitive vs standards?
- ▶ **Is it slow and inefficient?**
- ▶ Is it not standards-based?
- ▶ Is there hardware support?
- ▶ Are large companies investing in it?
- ▶ Is it difficult to transition to it?

Desktop

I+P frames, Full model
Titan V (4 years old)

	Encode	Decode
 VGA	47 FPS	91 FPS
720p	19 FPS	35 FPS
1080p	10 FPS	18 FPS

Decoding on Phones

720p, I+P frames
Slim Multi-bitrate model (not full version)

 Qualcomm™ snapdragon	30ms on Snapdragon 865 (2 years old)
 A12 Bionic	25ms on A12 (3 years old) 11ms on A14 (1.5 year old)

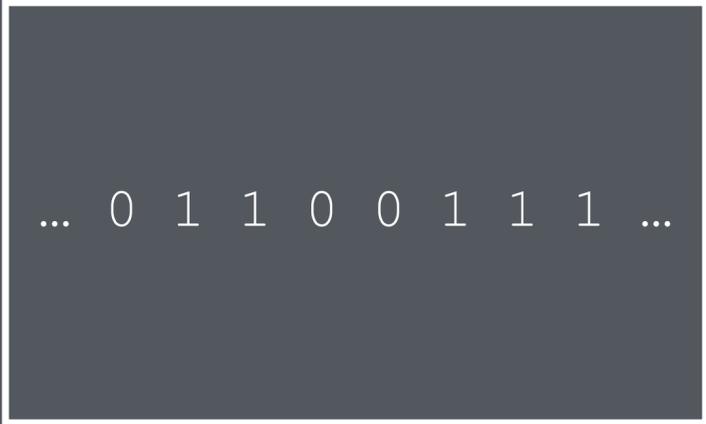
Questions about AI-Native compression



- ▶ Is it competitive vs standards?
- ▶ Is it slow and inefficient?
- ▶ **Is it not standards-based?**
- ▶ Is there hardware support?
- ▶ Are large companies investing in it?
- ▶ Is it difficult to transition to it?



Codec specification is standards-based



The encoded file is MP4 of raw bitstream (no structure)

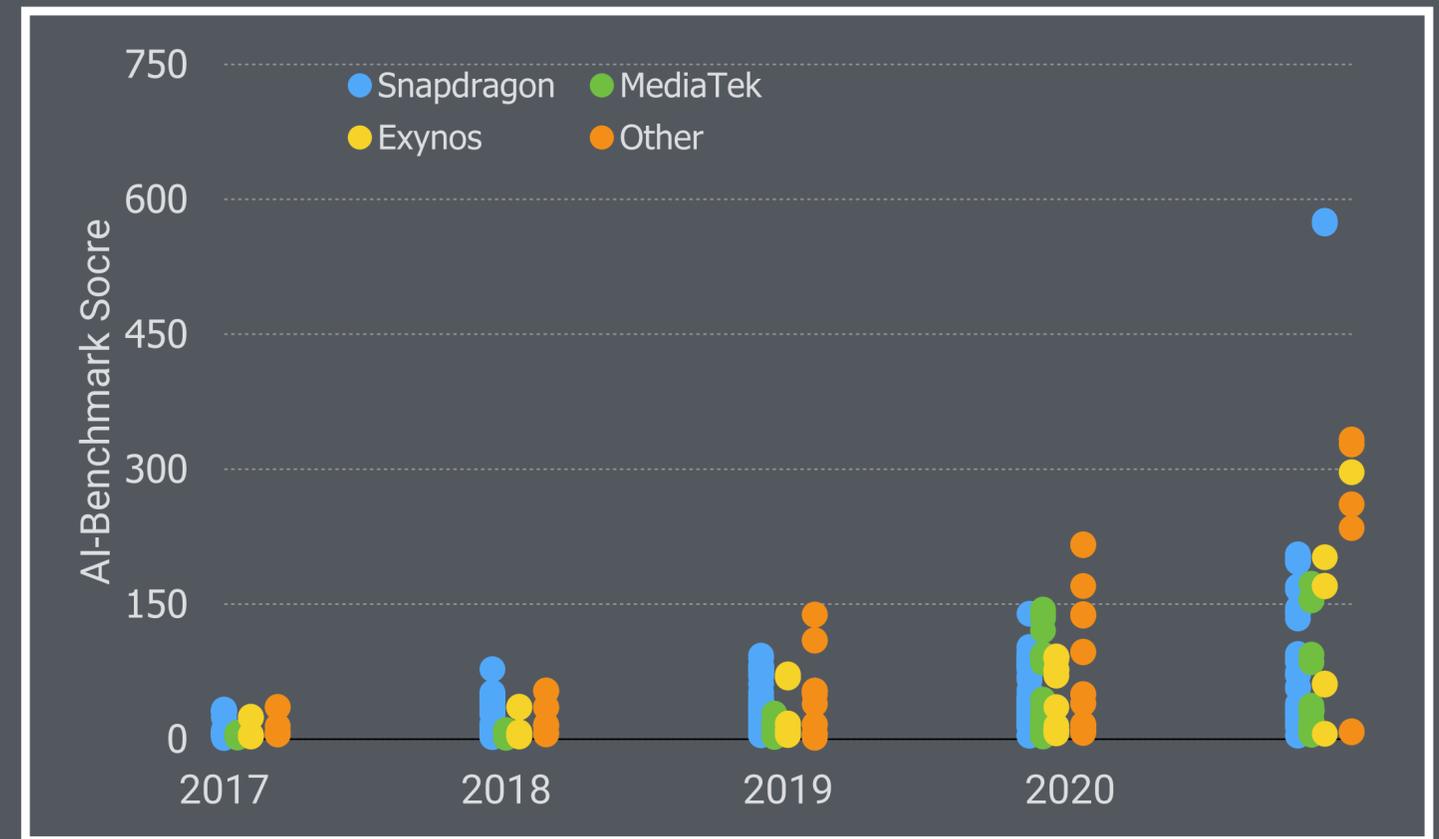
		iPhone	Pixel Phone	Safari	Chrome	Samsung TV	Roku	Fire TV
AV1	4K60	✗	✓	✗	✓	—	✗	✗
	1080P	✗	✓	✗	✓	—	✗	✗
H.265	4K60	—	✗	—	✗	—	—	—
	1080P	✓	✗	✓	✗	✓	✓	—
H.264	4K60	—	—	—	—	✓	✗	—
	1080P	✓	✓	✓	✓	✓	✓	✓

✓ = Yes
 ✗ = No
 — = Varies by model past 3 years

There are competing video standards

Questions about AI-Native compression

- ▶ Is it competitive vs standards?
- ▶ Is it slow and inefficient?
- ▶ Is it not standards-based?
- ▶ **Is there hardware support?**
- ▶ Are large companies investing in it?
- ▶ Is it difficult to transition to it?



DNN acceleration in phones grows dramatically

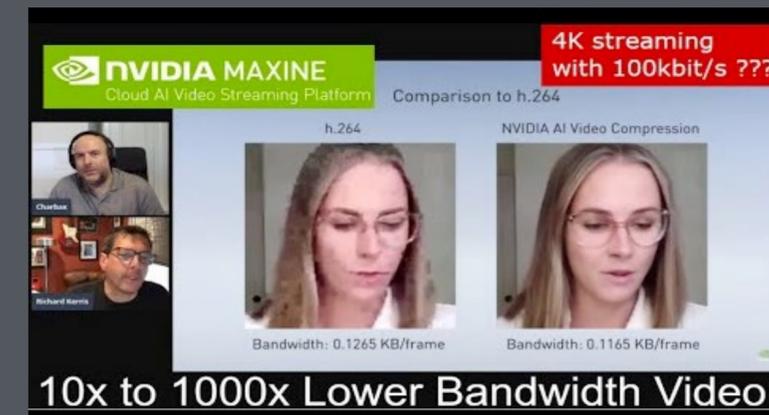
Questions about AI-Native compression

- ▶ Is it competitive vs standards?
- ▶ Is it slow and inefficient?
- ▶ Is it not standards-based?
- ▶ Is there hardware support?
- ▶ **Are large companies investing in it?**
- ▶ Is it difficult to transition to it?

Qualcomm



Nvidia Maxine



Google Starline

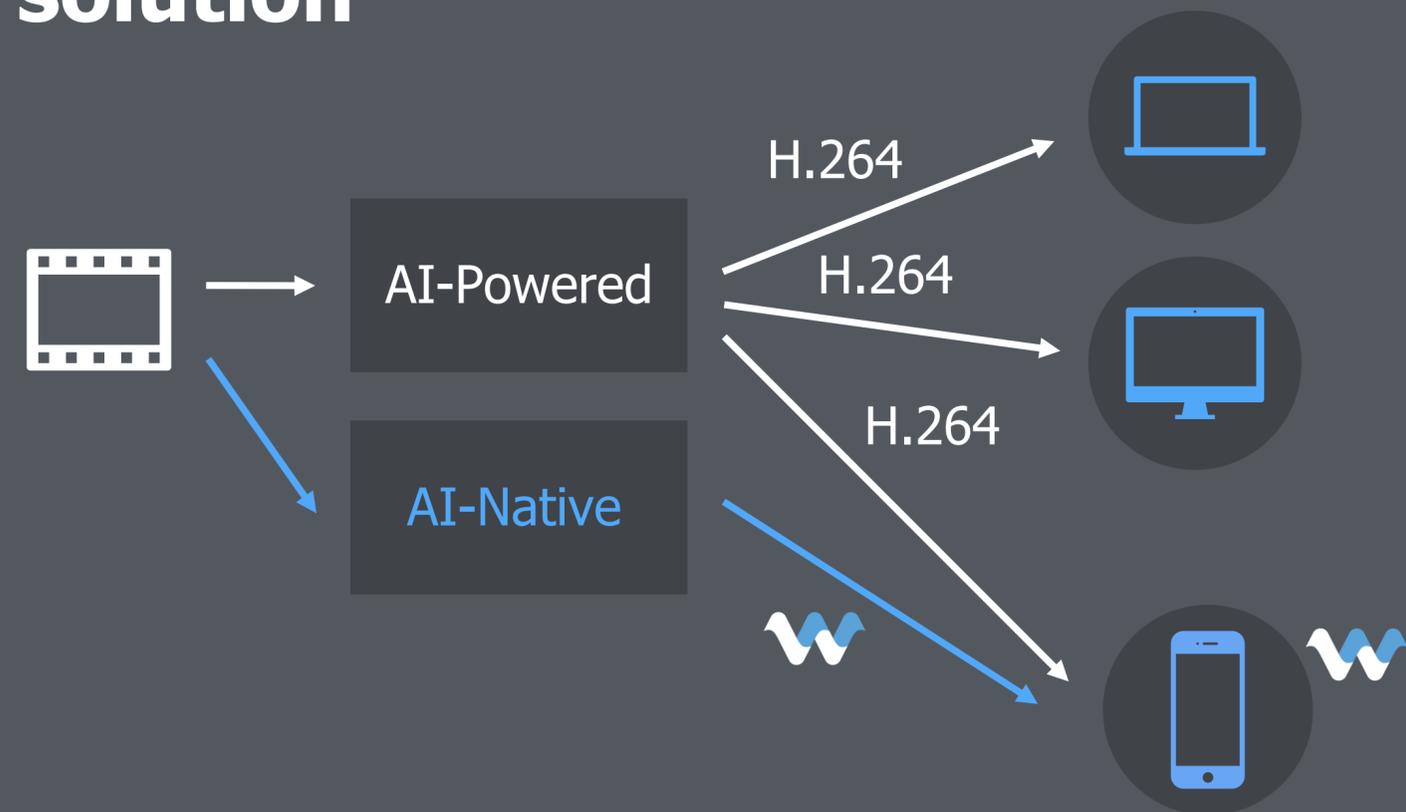


... and many more

Questions about AI-Native compression

- ▶ Is it competitive vs standards?
- ▶ Is it slow and inefficient?
- ▶ Is it not standards-based?
- ▶ Is there hardware support?
- ▶ Are large companies investing in it?
- ▶ **Is it difficult to transition to it?**

WaveOne will soon release VPaaS solution



Phase 1: H.264 compliant with ML
Phase 2: AI-Native to small but growing set of devices

Are we there yet?



WaveOne

We will get there sooner than you think!

s

Contact me at:

lubomir@wave.one