SVT-AV1-PSY

Presented by Julio Barba & Gianni Rosato

Introduction

- SVT-AV1-PSY = fork of SVT-AV1 with perceptual + quality-of-life features
- We'll explain:
  a. Why we began working on it
  b. Implementation details
  c. Current state of the project

Why?

- **SVT-AV1**: developed by Intel, Netflix, Meta *(now, mostly Meta and Intel)*
- **x265**: developed by MulticoreWare, Inc.
- **x264**: developed by VideoLAN & the community
- Between x264 & SVT-AV1, development shifted: perceptual focus → metrics focus
- Community wasn't satisfied with AV1 encoders
  - **libaom:** flexible, but very slow (and "benchmaxxed")
  - **SVT-AV1**: fast & great at metrics, still benchmaxxed; targeting arithmetic mean of PSNR/SSIM/VMAF
  - **rav1e**: less development momentum compared to SVT-AV1
  - **libaom forks**: not well-maintained, dubious value
- If nothing changed, the best open AV1 encoder wouldn't have many psychovisual features

Our Idea

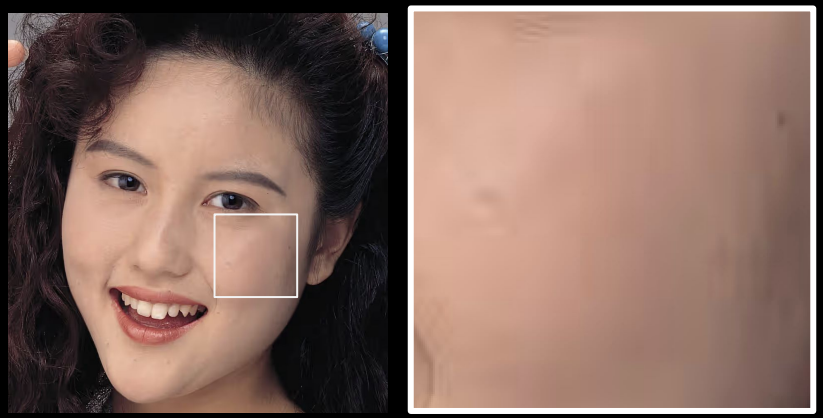
- State of libaom forks was unfortunate
  - Messy commit history
  - Inscrutable release cycle (sometimes sending patches over Discord!)
  - No relationship with Google
- Bring back community encoder development
  - Recruit devs from libaom forks, work together
  - Engage the community with new features
  - Simple, mainline-compatible release cycle
  - Ergonomic features, easy to use and understand
  - Active public presence → eventually, relationship w/ mainline team
- Forked SVT-AV1, called it SVT-AV1-PSY
  - Named after the first libaom fork (aom-av1-psy)
- Started work in Jan 2024 – Gianni started the project with BlueSwordM & Clybius, Julio came on soon after

Features

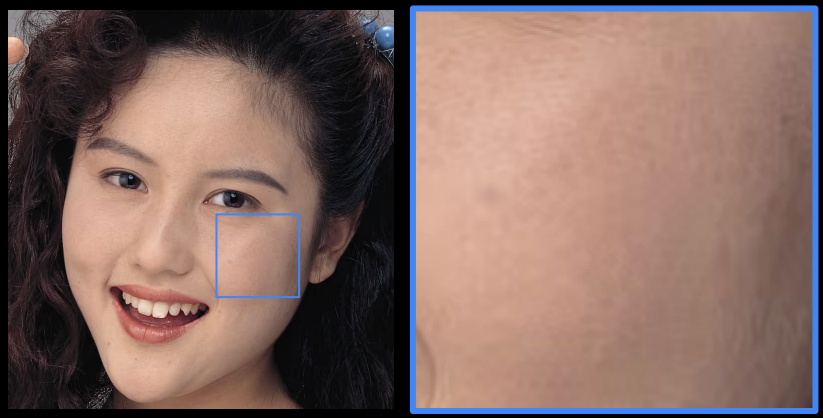
- **13** releases since launch, from Feb 2024 to Apr 2025
- **20+** impactful features, **388** GitHub ☆'s
- A fun website (svt-av1-psy.com)
- Features we'll talk about today:
  a. Variance adaptive quantization ("Variance Boost")
  b. Extended and quarter-step CRF
  c. AC Bias
  d. Tune 4 "Still Picture"

Variance Boost

- Variance AQ implementation, adapted to use AV1's delta-q signaling
  - Initially an SVT-AV1 patch, then it became SVT-AV1-PSY's first feature
- Helps retain faint textures (e.g. skin) and improves low-contrast scenes
- Can be used together with other delta-q modes (e.g. SVT-AV1's TPL)
- Optimized for various kinds of content (e.g. SDR, HDR video, still picture)
- Always uses 64×64 superblocks
  - Finer quantizer granularity control > total signaling overhead



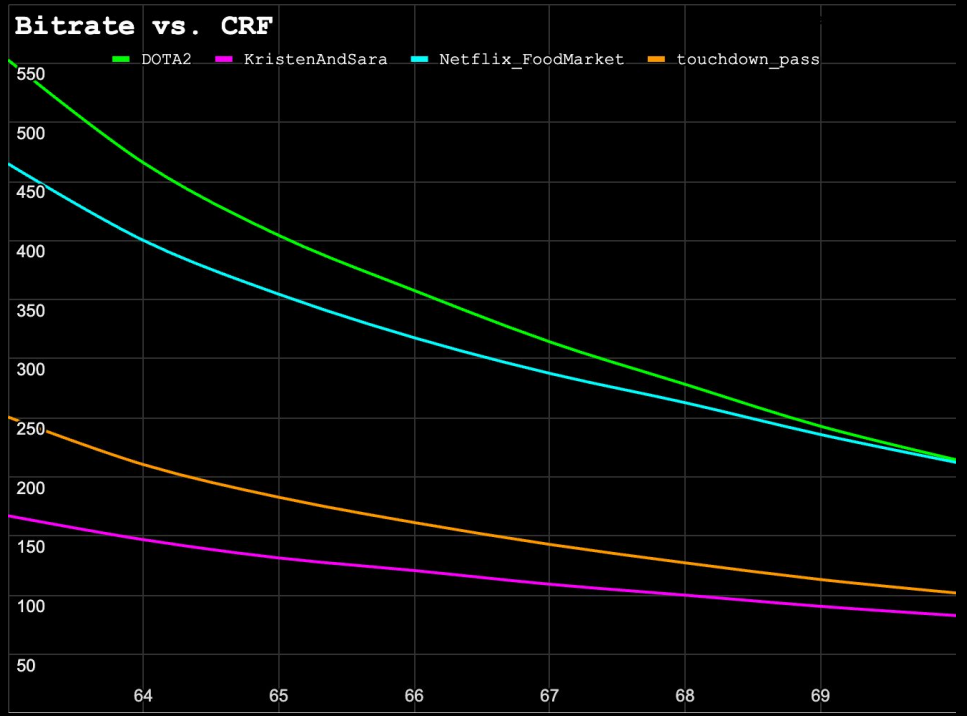
Variance Boost off



Variance Boost on

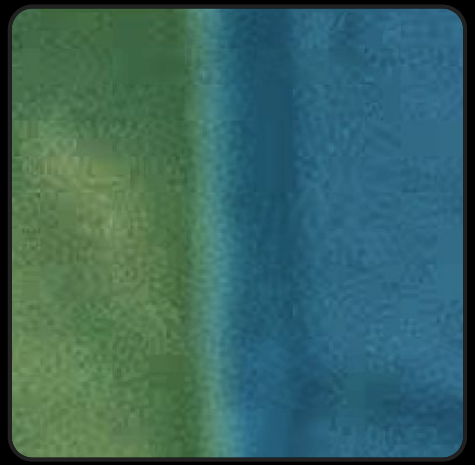
Extended and quarter-step CRF

- **Issue**
  - 64 QPs lack granularity (~7% size difference between consecutive values)
  - Highest CRF (63) can result in too large files in some cases
- **Solution**
  - Extend CRF [0-63] to [0.0-70.0] in 0.25 increments
  - New granularity: 0.25 QP ~= 1 qindex
  - CRF >63: compress QP scaling and bias RDO toward increased bit saving
- Achieves MPEG codec bitrate range versatility

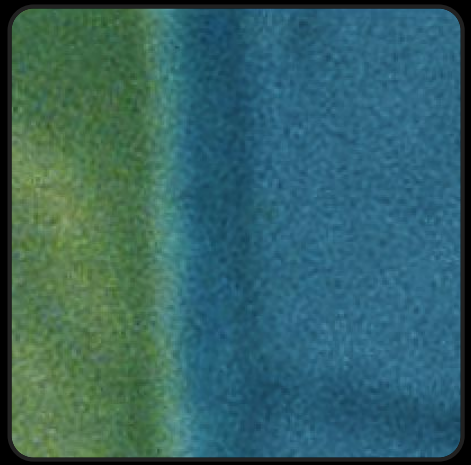


AC Bias

- Energy-preserving in-loop metric
- Similar to x264 and x265's "psychovisual rate distortion" (PSY-RD)
  - Lower strengths help retain sharpness and acuity of textures and scenes with complex motion
  - Higher strengths can improve film grain retention significantly



AC Bias strength 0.0



AC Bias strength 4.0

AC Bias: Intuition

- Walsh-Hadamard transform to go from pixel space → frequency space
- Comparing source pixels to recon after transforming = SATD (Sum of Absolute Transformed Differences)
- Comparing pixels normally = SAD (Sum of Absolute Differences)
- Hadamard transform still has a DC component → subtract SAD to account for this
- So: SATD - SAD = energy distortion
- Biases in-loop distortion toward visual energy
- Additional considerations:
  - Weights for different temporal layers
  - Calibrated so that strength = 1.0 makes sense for most users
  - Enabled by default

Tune 4 “Still Picture”

- AV1 isn’t just a video codec; it is also an image codec!
- The approach at the time was build a good video encoder → good images will come out
- We can test this: images have good perceptual metrics, like SSIMULACRA2 & Butteraugli
- Tune 4 is five smaller features in one:
  - Quantization matrix scaling curve
  - DLF sharpness
  - New Variance Boost curve
  - Screen content detection improvements
  - RDO Lambda tweaks
- All Tune 4 features were optimized around the CID22 Validation Set, and measured against other image datasets later

Stock

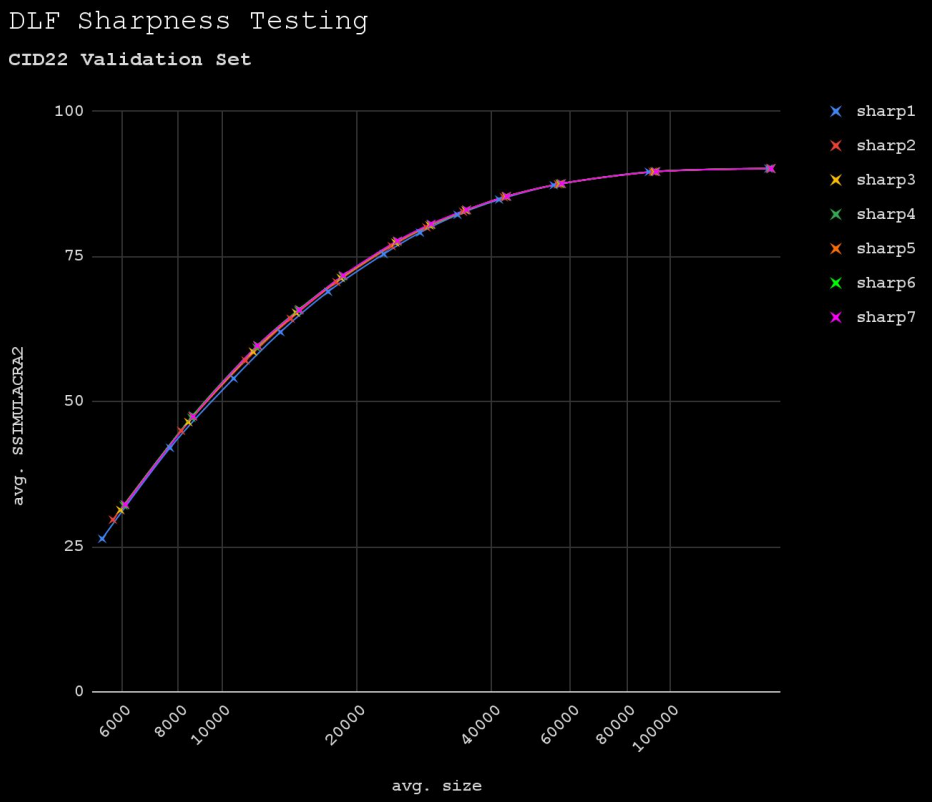
- By default our features helped, and the metric rewards them
- What else can we do?

Quantization Matrix Scaling

- AV1 can specify different levels of quantization to frequency components
  - AV1 includes 15 predefined QMs
- RD curve at each QM, pick the best at each section of the curve
- Result: better multi-scale detail preservation
  - Complex textures, camera noise

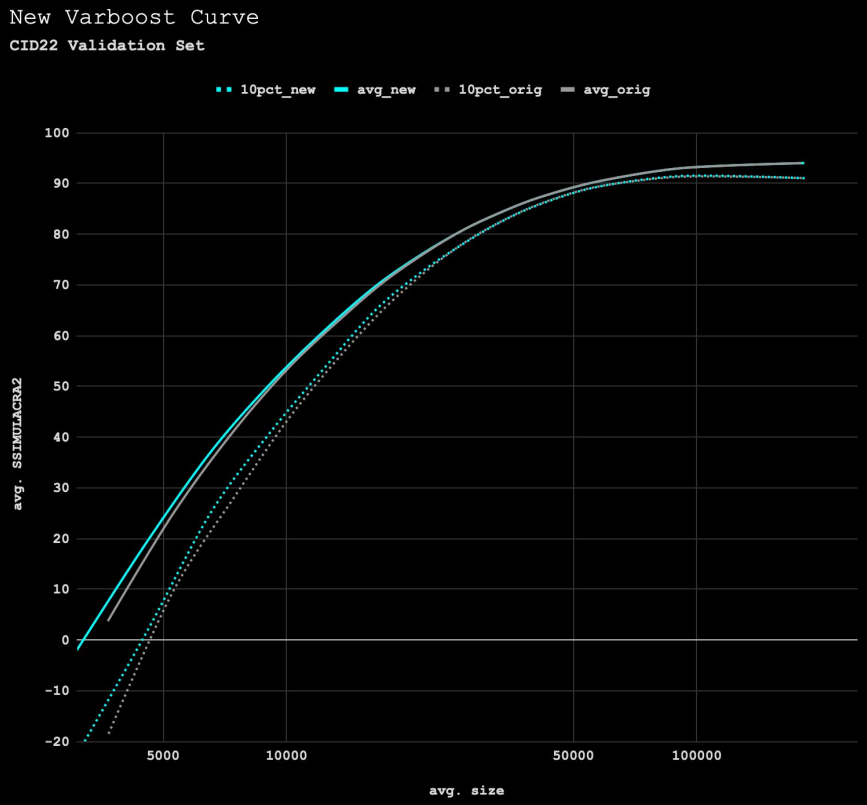
DLF Sharpness

- Deblocking loop filter controls how the boundaries between blocks are handled
- SVT-AV1-PSY has sharpness controls for this; try each one & plot
- Smoother DLF might help video by improving inter-frame consistency; images are clearly a different story



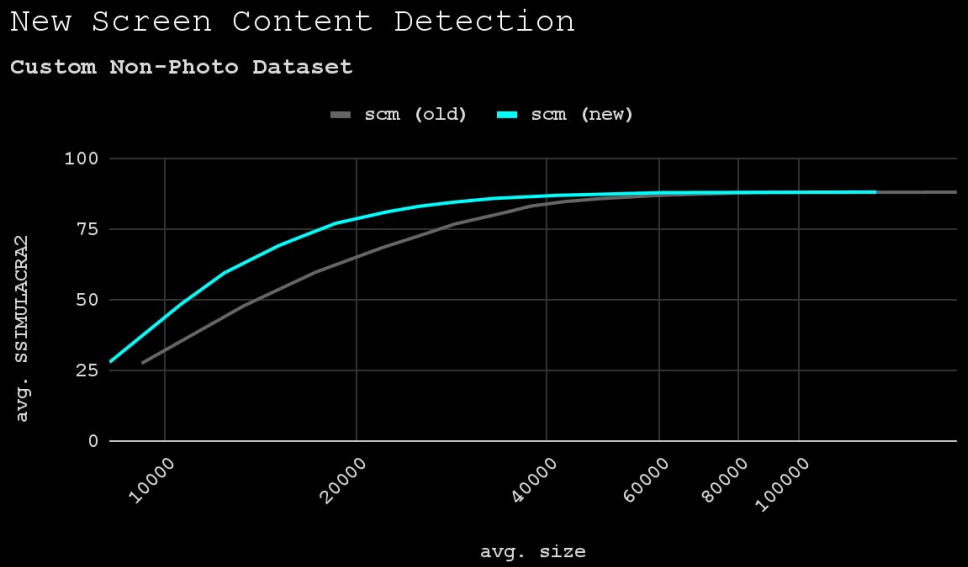
New Variance Boost Curve

- Don’t boost low variance as eagerly
- Taper back high Variance Boost more gradually
- Helped with 10% lows, which helps encoder consistency across images
- Subtle improvements in intra-image consistency, texture retention



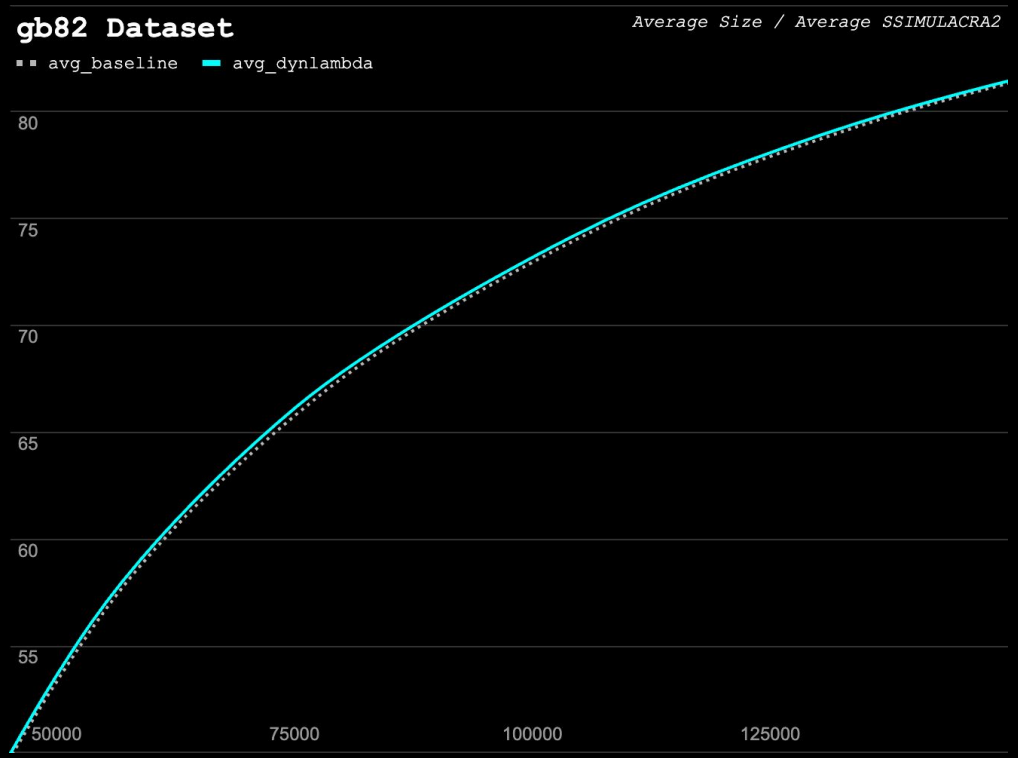
Anti-aliasing Aware Screen Content Detection

- Old algorithm (inherited from libaom) wasn't very effective on images
  - Naive algorithm based on total color counting 16×16 regions
  - Bad at detecting screenshots with anti-aliased fonts, a.k.a. most screenshots
- New algorithm is anti-aliasing aware; detection rate increased



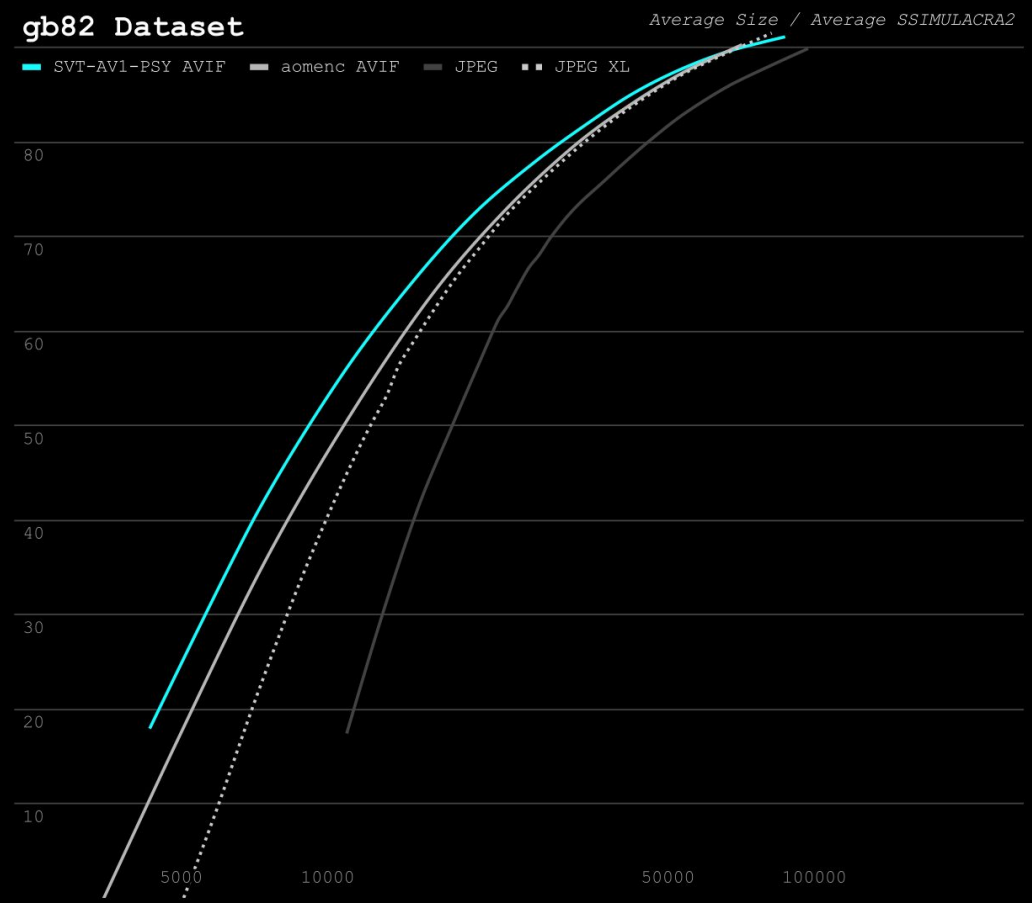
RDO Lambda tweaks

- Lambda = Lagrange multiplier in rate distortion
  - Cost = Distortion + λ * Rate
  - High λ = rate reduction
  - Low λ = distortion reduction
- For Still Picture tune:
  - At higher & lower quantization, decrease lambda
  - In between, increase
- Reduced file sizes, preserved subjective quality → slightly better efficiency



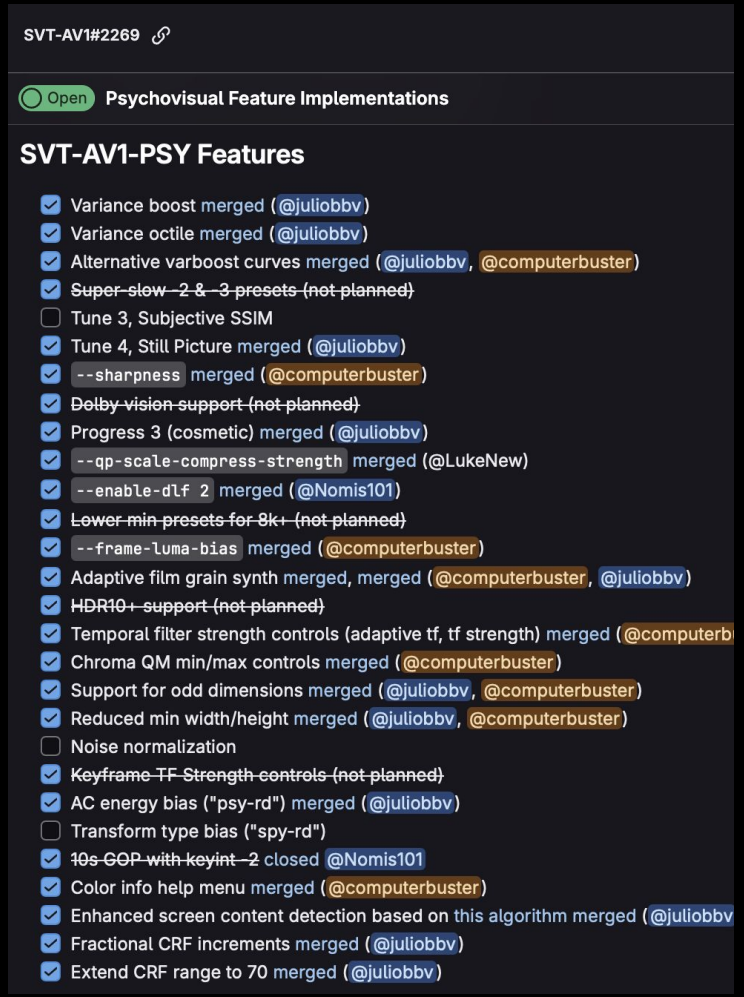
Tune 4 Final Results

- We are ahead of aomenc on its best performing settings, as well as JPEG XL
- Perceptually, results look just as good
- Butteraugli and DSSIM also look good
- This was state of the art
- Eventually merged to SVT-AV1 as Tune 3 “Tune IQ” (Image Quality)
  - Also made it to libaom as “Tune IQ” – thanks Julio!



SVT-AV1-PSY mainlining

- Work to merge PSY code to upstream SVT-AV1
  - Wave 1: Gianni (@*computerbuster*), Dec 2024
  - Wave 2: Julio (@*juliobbv*), Sept-Oct 2025
  - A few features were upstreamed by other community members
- We discussed a contribution plan with the SVT-AV1 team
  - Interest, collaboration and feedback have been outstanding!
- Issue (!2269): gitlab.com/AOMediaCodec/SVT-AV1/-/issues/2269



Other notable PSY features → SVT-AV1

- QP scaling compression
- Frame luma bias
- Adaptive film grain
- Max transform size
- Odd resolutions, <64×64 inputs
- Dynamic delta-q res switching (1, 2, 4, 8) support
- Detailed (and colorful) progress report
- And many more!

```
Svt[info]: SVT [config]: main profile   tier (auto)     level (auto)
Svt[info]: SVT [config]: width / height / fps numerator / fps denominator              : 1920 / 1080 / 24000 / 1001
Svt[info]: SVT [config]: bit-depth / color format                                      : 8 / YUV420
Svt[info]: SVT [config]: preset / tune / pred struct                                   : 3 / PSNR / random access
Svt[info]: SVT [config]: gop size / mini-gop size / key-frame type                     : 240 / 32 / key frame
Svt[info]: SVT [config]: BRC mode / rate factor                                        : CRF / 40
Svt[info]: SVT [config]: AQ mode / Variance Boost strength / octile / curve            : 2 / 2 / 5 / 0
Svt[info]: SVT [config]: sharpness / luminance-based QP bias                           : 0 / 0
Svt[info]: SVT [config]: QP scale compress strength                                    : 1
Svt[info]: -----------------------------------------
Encoding:  145/241 Frames @ 8.54 fps | 975.70 kb/s | Size: 1.23 MB [2.04 MB] | Time: 0:00:17 [-0:00:11]
```

libaom improvements

- The entire Tune IQ feature set was ported over to libaom
  - `avifenc -a c:tune=iq input.png output.avif`
  - Work is ongoing to make Tune IQ default in libavif (ETA Jan 2026)
- Some improvements that came **from** libaom were later ported **to** SVT-AV1
  - Adaptive DLF sharpness
  - Fast anti-aliasing aware screen content detection mode
  - Intra Block Copy block hashing perf improvements
- Work on Tune IQ helped improve AV2 coding tools for future still picture tuning
  - CWG-D130: Quantization Matrices Coding for AV2
  - CWG-F089: Delta-q Entropy Coding Optimization

Thank you!

git shortlog -sn 16273d98..HEAD
Gianni
BlueSwordM
Julio Barba
Clybius
BillyBob1354
JND94
bex39
kremzli
quietvoid