

# Metric Blind-Spot Field Guide

Where PSNR, SSIM and VMAF lie — and what to measure instead. Keep it beside your quality dashboard.

Content / case	Why the score lies	Metrics hit	What to measure instead
<b>Film grain / texture</b>	Random detail never lines up pixel-for-pixel; AV1 grain synthesis mismatches grain positions.	<b>PSNR, SSIM, VMAF</b>	Disable grain synthesis when measuring; measure the de-grained signal; confirm by eye.
<b>High motion</b>	Eye masks detail during motion; VMAF v0's motion feature is unbounded → overprediction.	<b>VMAF v0, PSNR</b>	Use the current VMAF model; keep frame rate fixed; spot-check the fastest scenes.
<b>High frame rate (60 fps)</b>	Consecutive-frame motion looks small, so v0 underpredicts vs 24/30 fps.	<b>VMAF v0</b>	Never compare across frame rates; use VMAF v1; always report fps.
<b>Dark scenes / HDR</b>	Few bright pixels to 'disagree'; no deployed public HDR VMAF model yet.	<b>PSNR, SSIM, VMAF</b>	Inspect shadows; weight subjective checks to dark scenes; treat HDR numbers as indicative.
<b>Banding (skies, fades)</b>	A smooth ramp becomes flat plateaus, which barely moves pixel statistics.	<b>PSNR, SSIM, VMAF v0</b>	Run CAMBI alongside VMAF; gate on banding separately.
<b>Chroma / color</b>	Luma-only computation ignores the color channels entirely.	<b>luma PSNR, SSIM, VMAF v0</b>	Measure chroma (U/V) or a chroma-aware model; check saturated regions.
<b>Screen content, text, gaming</b>	Trained on natural video; flat regions, hard text edges and patterns differ.	<b>PSNR, SSIM, VMAF</b>	Spot-check legibility; run a subjective test on the real content.
<b>Animation / flat synthetic</b>	Domain mismatch with photographic training; flat regions also band.	<b>PSNR, SSIM, VMAF</b>	Add a small subjective check; watch flat regions for banding.
<b>Enhancement (sharpening)</b>	The metric rewards looking different, not fidelity, so it is gameable.	<b>VMAF (standard)</b>	Check the gain against VMAF-NEG; if it vanishes, it was enhancement.

## Before you trust a score

Read a panel of metrics, not one · pool for the worst case · measure on your own content · keep every comparison apples-to-apples (same resolution, frames, reference, model, pooling) · confirm the hard frames by eye. A metric is a model with a domain; outside it, the number lies confidently.