

如何正确的评估视频画质？

王伟(wangwei1237@gmail.com)

1. 什么是视频画质评估？为什么要做画质评估？
2. 影响视频画质的因素有哪些？
3. 提升视频画质必要性和演进路径是什么？
4. 视频评测的方法、标准、干扰因素有哪些？
5. 视频画质评测工作的感想：唐僧扫塔



科研第一步？
优秀的综述论文

1. 什么是视频画质评估？

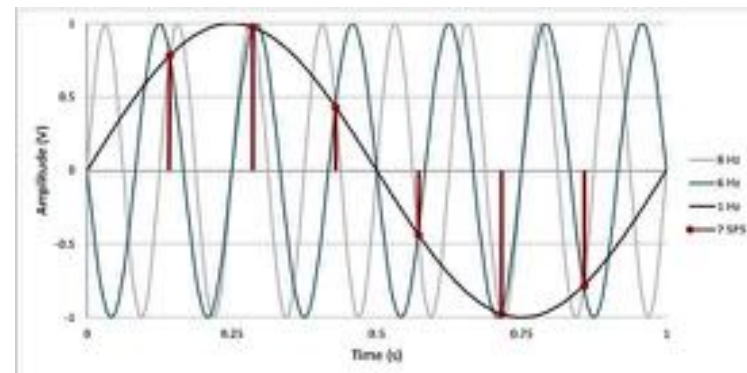
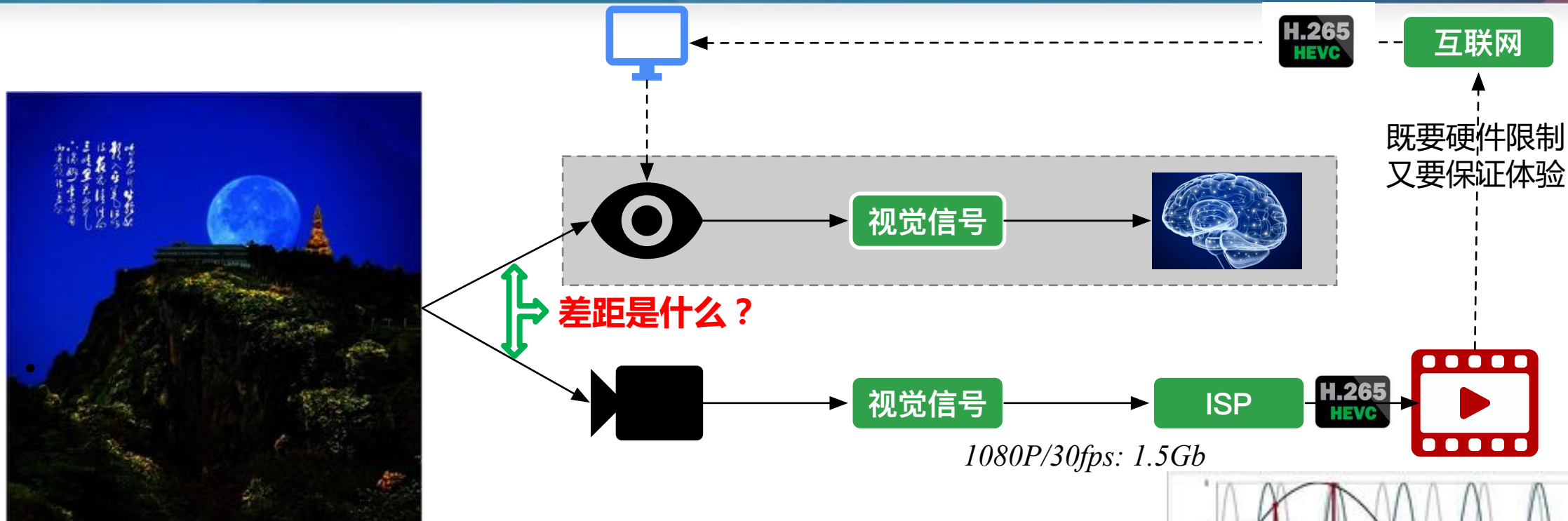
我们是如何感知到丰富多彩的世界的？

- 仗剑走天涯，透过眼睛看世界：世界那么大，我想去看看~
- 从早刷到黑，透过屏幕看世界：世界那么小，只在一屏间~



视频技术的发展目标：产生更真实的视觉体验。

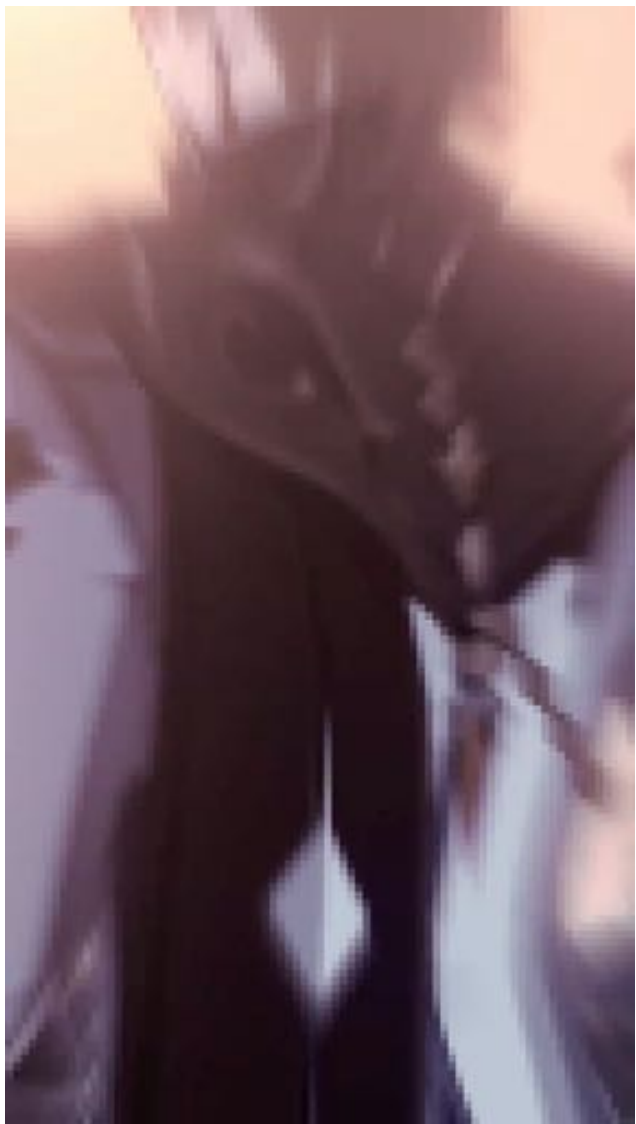
1. 什么是视频画质评估？



- 采集信号的能力如何，也就是摄像头镜头的能力如何？
- 从采集到最终人在屏幕上看到视频，这个流程的信息损失如何？
- 对于不同的视频处理系统，视频画质评估还用来评估不同系统的优劣？

2. 影响视频体验的因素有哪些？

1. 分辨率(144P->1080P)



2. 帧率



3. 亮度



草的RGB信息
(69,117,32)
(167,218,76)

当问题在某个空间
无法解释时怎么办？

空间转换

草的HSL信息
(94,57,29)
(82,65,57)

2. 影响视频体验的因素有哪些？

4. 位深(8bit->10bit)



8-bit

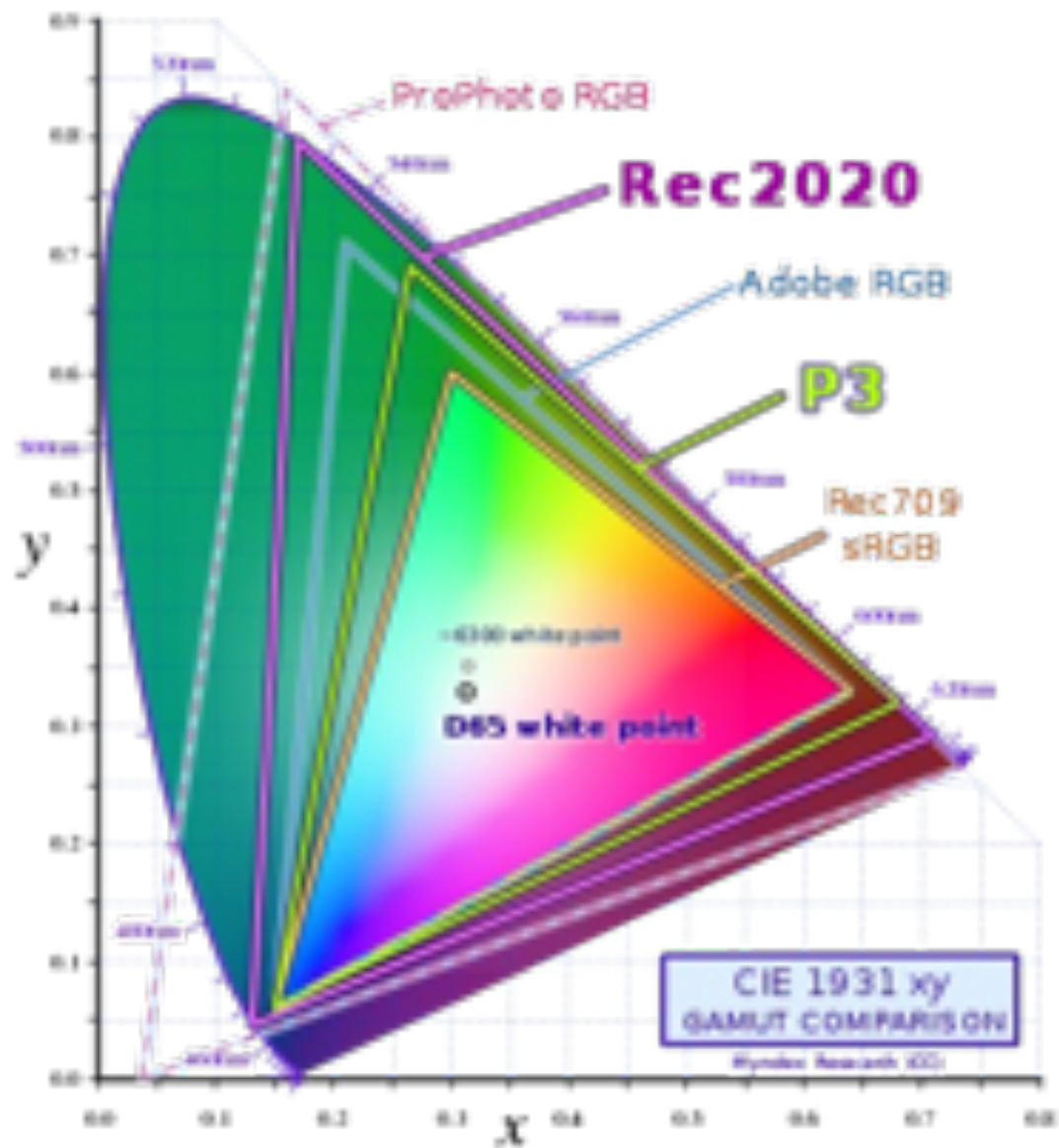


10-bit

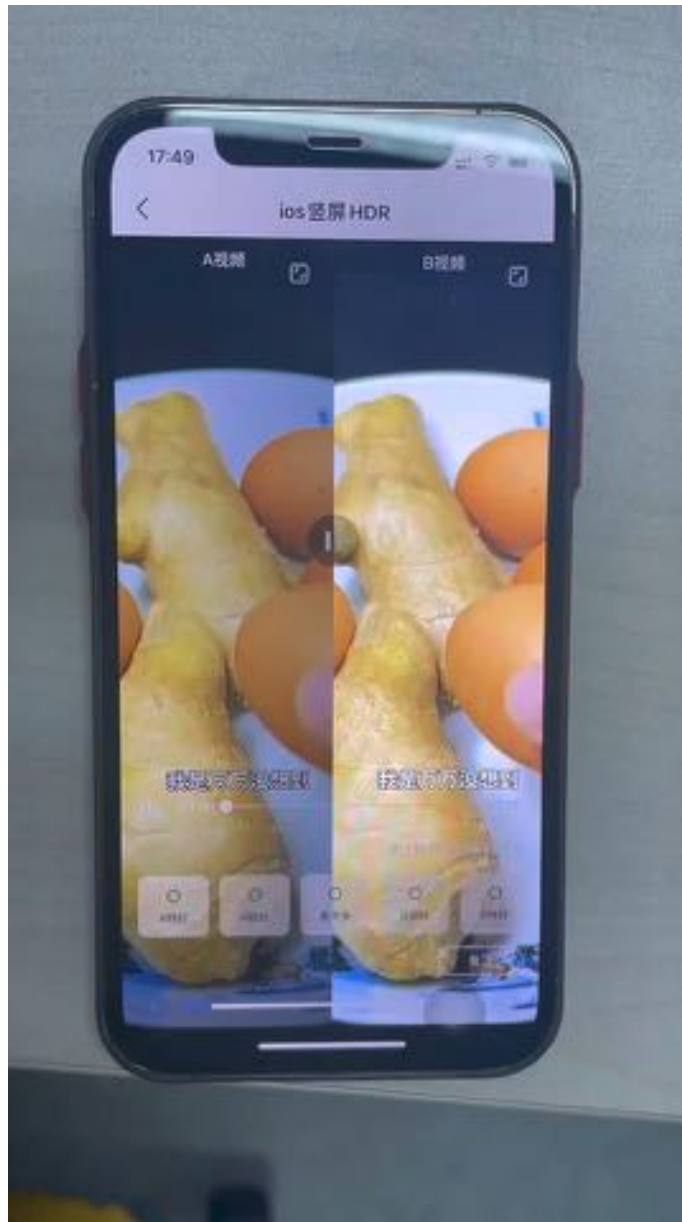


2. 影响视频体验的因素有哪些？

5. 色域(bt709, bt2020)



SDR&HDR的对比区别



3. 为什么要不断的提升画质？

1. 整体看工业界和学术界，提升画质是不断追求更好视频体验的结果。

2. 消费级别的采集设备不断升级。

3. 消费级别的显示设备不断升级。

HUAWEI Mate 40 Pro

屏幕	尺寸	色彩
	6.76 英寸	1670 万色, DCI-P3 广色域
	*显示屏采用圆角设计, 按照标准矩形测量时, 屏幕的对角线长度是 6.76 英寸 (实际可视区域略小)。	
	类型	分辨率
	OLED, 最高支持 90 Hz 刷新率、240 Hz 触控采样率。	FHD+ 2772 x 1344 像素

iPhone 13 Pro

摄像头

Pro 级 1200 万像素摄像头系统 (长焦、广角及超广角)
长焦: $f/2.8$ 光圈
广角: $f/1.5$ 光圈
超广角: $f/1.8$ 光圈和 120° 视角

2K屏幕播放360P视频



4. 不断增加的竞对产品。

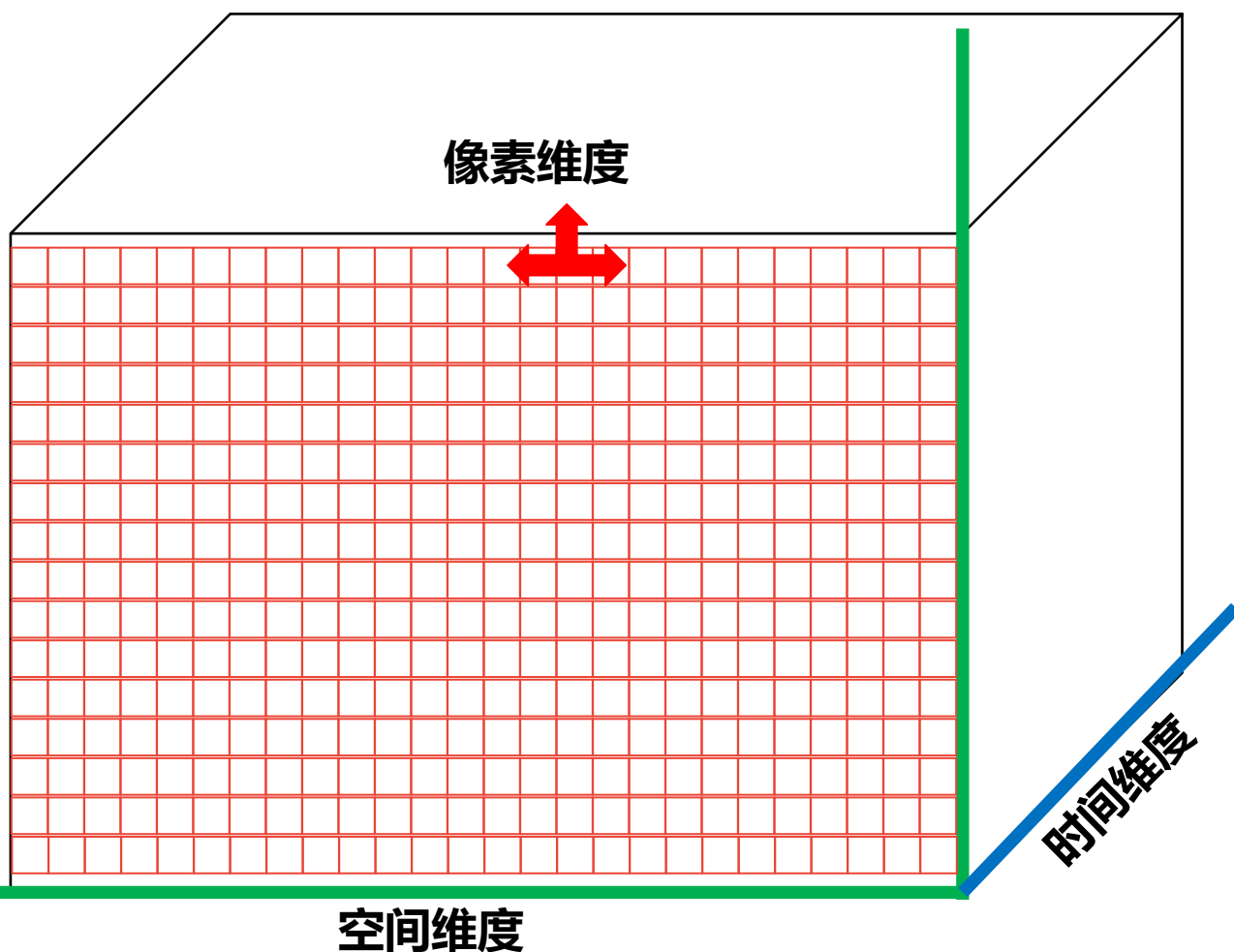
- 1. 视频画质的基线水平取决于行业中的竞对产品，并且画质水平是各家产品对外宣传的有效手段。
爱奇艺：帧绮映画；B站：4K/120fps；西瓜：4K高清修复技术；快手：质臻影音
- 2. 考试时，自己的成绩排名不取决于自己考的有多好，而取决于其他人考的怎么样。
- 3. 逆水行舟，不进则退。

3. 为什么要不断的提升画质？



4. 提升画质的演进路径是什么？

数字视频其实就是对连续信号的采样：空间采样，时间采样，像素采样



空间维度：像素个数不断增加，分辨率不断提升

- 360P
- 540P
- 720P
- 1080P
- 4K/8K

时间维度：每秒视频帧数不断增加，帧率不断提升

- 20 fps
- 60 fps
- 120 fps

像素维度：单个像素的质量不断提升

- SDR
- HDR

5. 视频画质评测的方法有哪些？

视频/图像质量评价(Video/Image Quality Assessment)：

- 通过主、客观的方式对两个主体内容相同的视频/图像信息的变化与失真进行感知、衡量与评价。
- 评价方法分为**主观评价方法**和**客观评价方法**两种。

主观评估

基于人类的主观评估来确定图像或视频内容的实际感知质量，也就是通过观察者的评分来衡量。

比如从用户体验的角度来观察视频的体验质量，也就是目前比较🔥的QoE (*quality of experience*) 。

优点：更接近用户的感知

缺点：不可重复执行，资源消耗大，结论受各种因素的影响比较大。

客观评估

计算机根据算法计算出视频/图像的质量指标，评估信号的保真度，或者评估图像之间的相似性

比如计算图像之间差异的指标峰值信噪比(PSNR)；将图像质量下降转化为感知结构信息变化的指标结构相似度(SSIM)；VMAF等

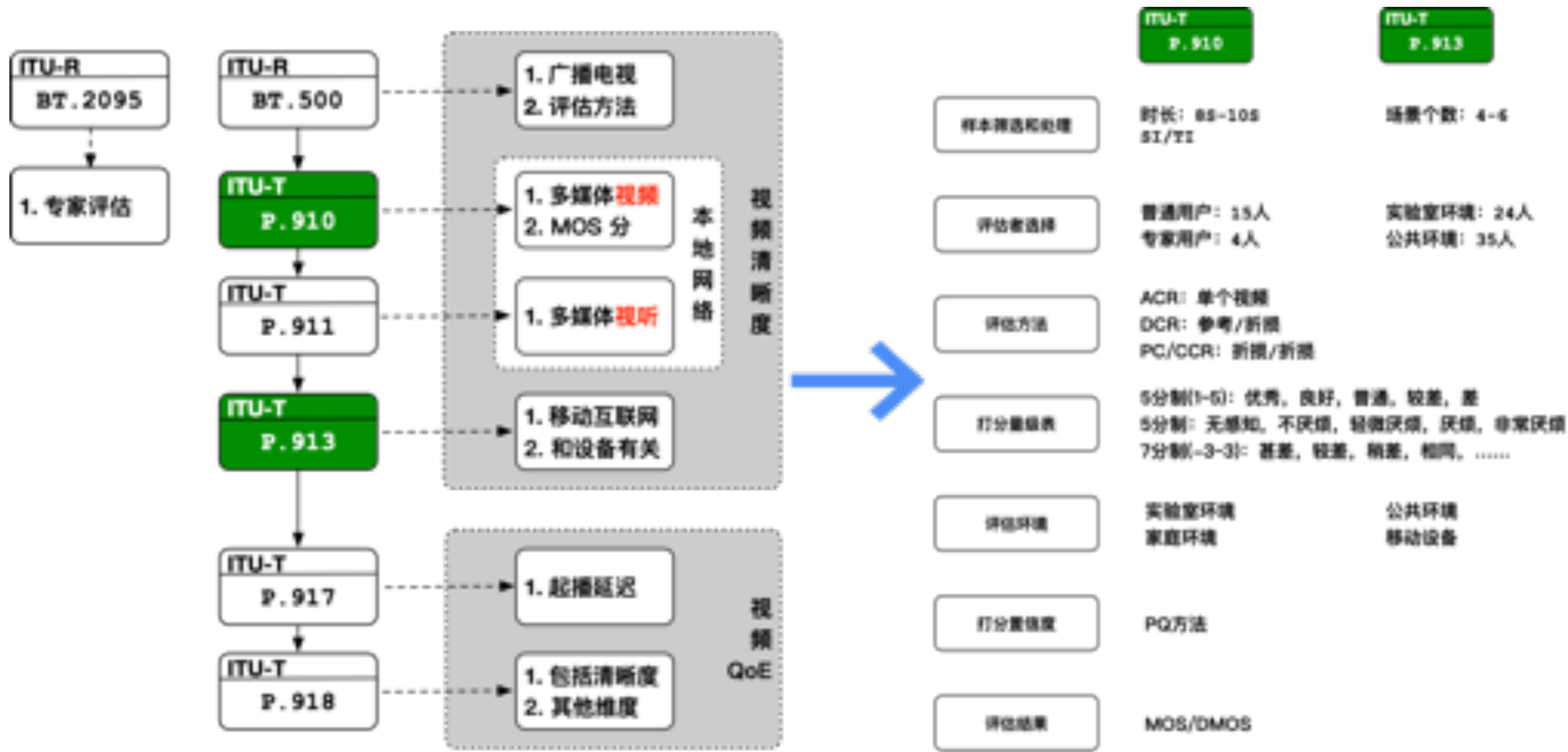
优点：可重复执行，快速高效

缺点：不一定和用户的主观感受一致

[MSU 视频编解码大赛](#)的比赛规则也分为：主观评估和客观评估两个指标。

客观评估：衡量大盘数据分布；实际编解码器优化效果最终还要以主观评估为主。

6. 视频画质主观评测的标准有哪些？



7. 如何正确的评估视频画质？

有标准
找人打分呗



复杂的事情往往会用简单的表象迷惑我们！

- 悬链线：[项链形成的曲线是什么](#)？科学家花了近200年才被破解~
- 苹果为什么会掉到地上呢？



8 影响主观评测结果置信度的因素都有哪些？

有了国际标准，按照标准来执行就能得到正确的结论吗？从实践结果看，并非是肯定的。

- 用某一类视频（旅游、生活）来作为样本评估整体效果？
 - 评测样本的抽样是否合理，是否可以表征产品的整体？抽样数据如何代表整体？
- 评测时用到的工具是否能够代表用户的消费场景？
 - 视频是半屏评测还是全屏评测？半屏是16:9的视窗大小还是3:2的视窗大小？如何解决端上增强处理的问题？
- 选择哪些用户来对视频进行打分评估呢？
 - 不同用户打分不一致如何处理？如何判断用户打分是否置信？当某个用户打分严重偏离其他用户打分时，如何处理？
- 在数据分析的时候采用什么样的数据处理方法呢？
 - 平均值？众数？中位数？
 - 如何计算CodecA-CodecB之间的码率差异？所有样本取均值然后计算差异 or 每个样本计算差异然后取均值？
 - $(\sum B_i / \sum A_i) - 1$ or $(\frac{1}{N} \sum (B_i/A_i)) - 1$ 这两种方式之间的区别又是什么？
 - 如何处理辛普森悖论的问题？

8 影响主观评测结果置信度的因素都有哪些？

根据我们近2年的实践，我把影响视频主观评估置信度的因素归结为4个方面：

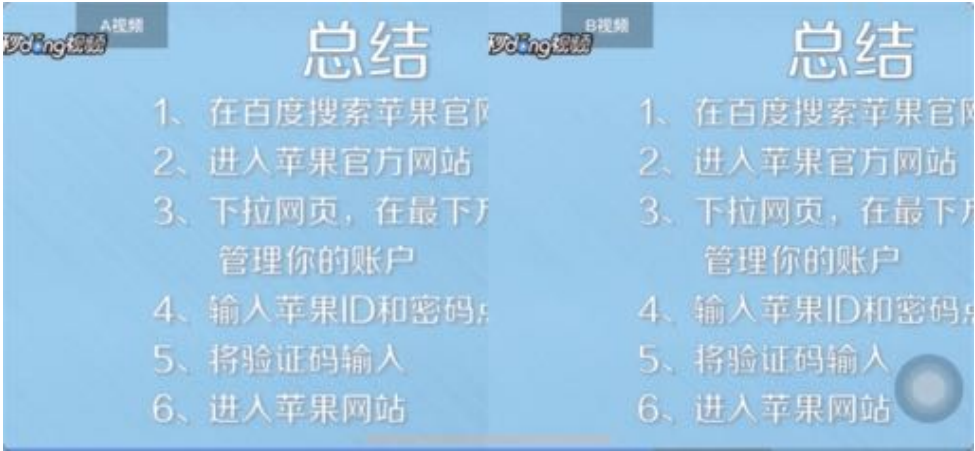
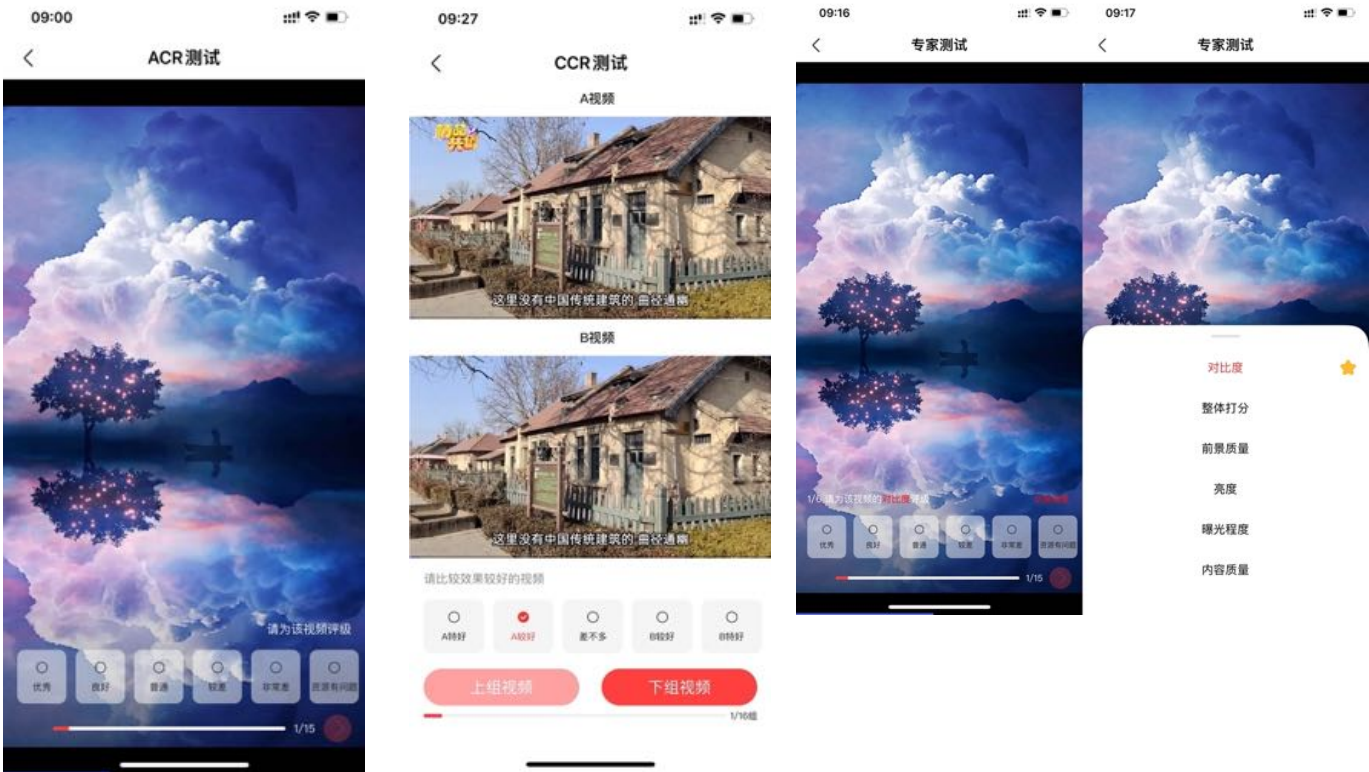
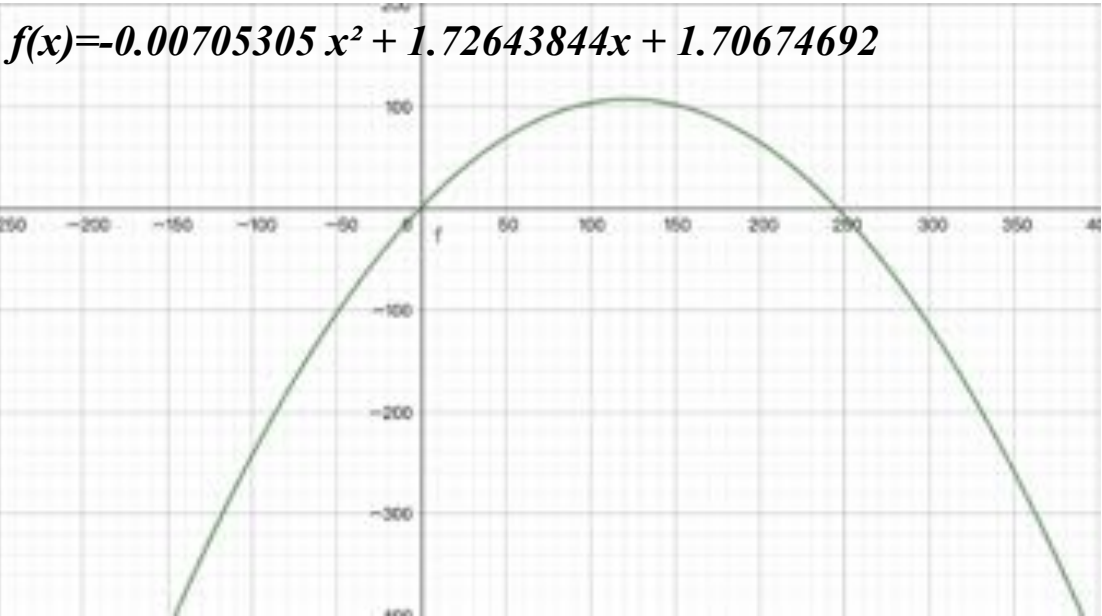
- 1. 评测工具**
- 2. 评测样本**
- 3. 评测用户**
- 4. 数据分析方法**

8 影响主观评测结果置信度的因素都有哪些？

自研评测工具（灵镜）：依托国际标准，针对业务使用场景，实现80+评估场景

- 1. 支持所有的业务场景：PC/Android/iOS
- 2. 支持所有的国际标准：ACR/DCR/CCR/专家评估
- 3. 特殊端的视频处理：华为NPU的视频超分技术
- 4. 针对不同设备提供不同样本集

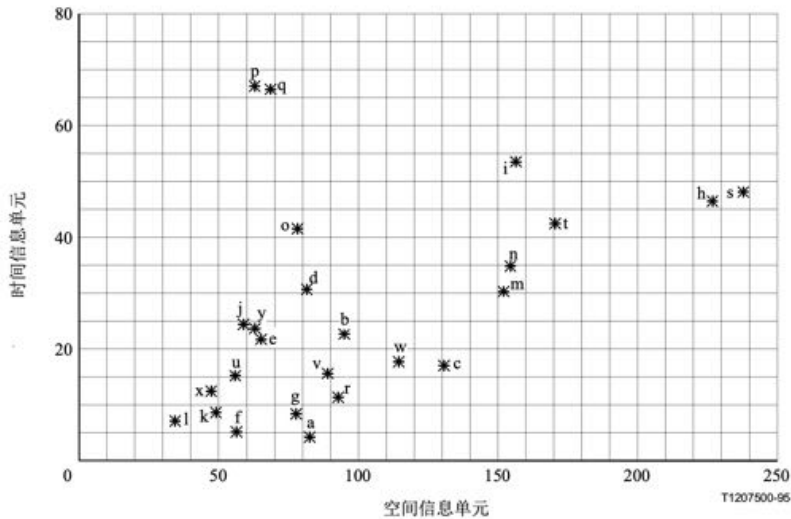
VMAF算法中利用电视模型对手机模型的打分拟合函数



8 影响主观评测结果置信度的因素都有哪些？

样本层面：让样本更具备代表性，消除样本带来的评测偏差。

1. 更全面的抽样：结合视频SITI、业务的视频分类数据等多维度抽样



2. 样本从完全透明-》单盲-》双盲



8 影响主观评测结果置信度的因素都有哪些？

评测去IOE化，制定标准，控制用户关注焦点；上报信息，区分用户的终端设备。

标准的视频评测实验室（IOE）：

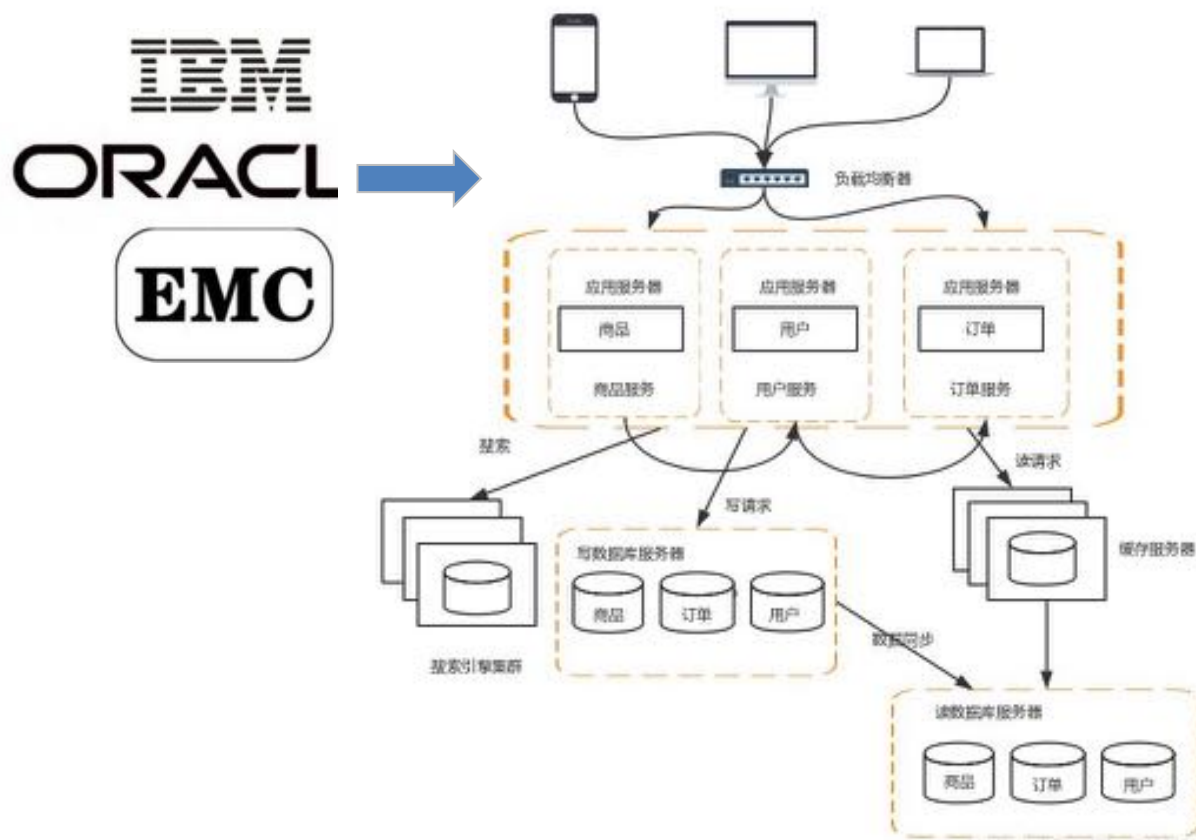
评测者评测质量较好，设备信息控制简单，但是效率比较低，成本比较大。



基于众测用户的评测方案（去IOE化）

评测者质量如何控制？用户设备如何控制？评测方式如何控制？

评测者打分置信度如何判断？



8 影响主观评测结果置信度的因素都有哪些？

如何科学的整理评测结果

1. 两个不同编解码器的主观全参考MOS打分

	codec A	codec B
720P	85	80
1080P	80	90

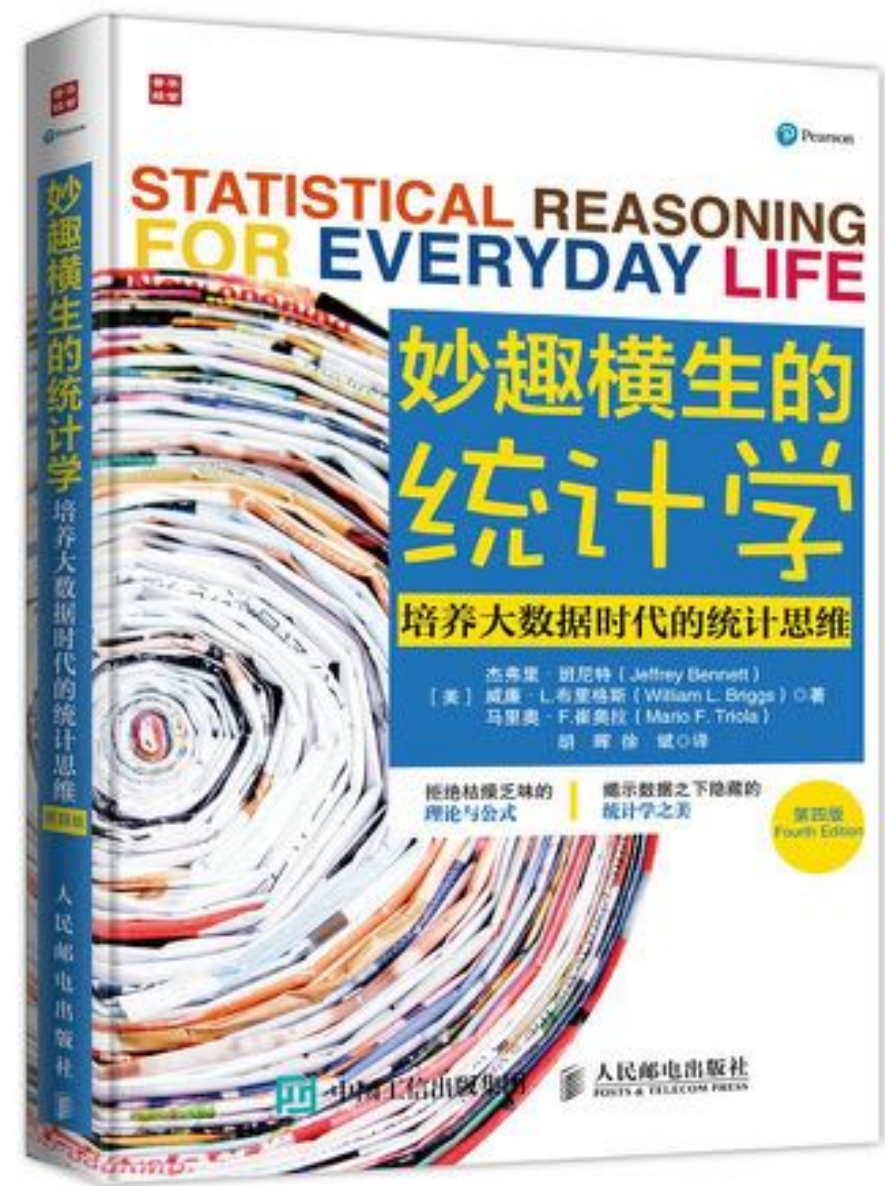
MOS打分范围[0,100]，得分越高越好

2. 根据如上的结果，哪个编解码器更好？

B好？因为B的平均MOS分是85分，高于A。

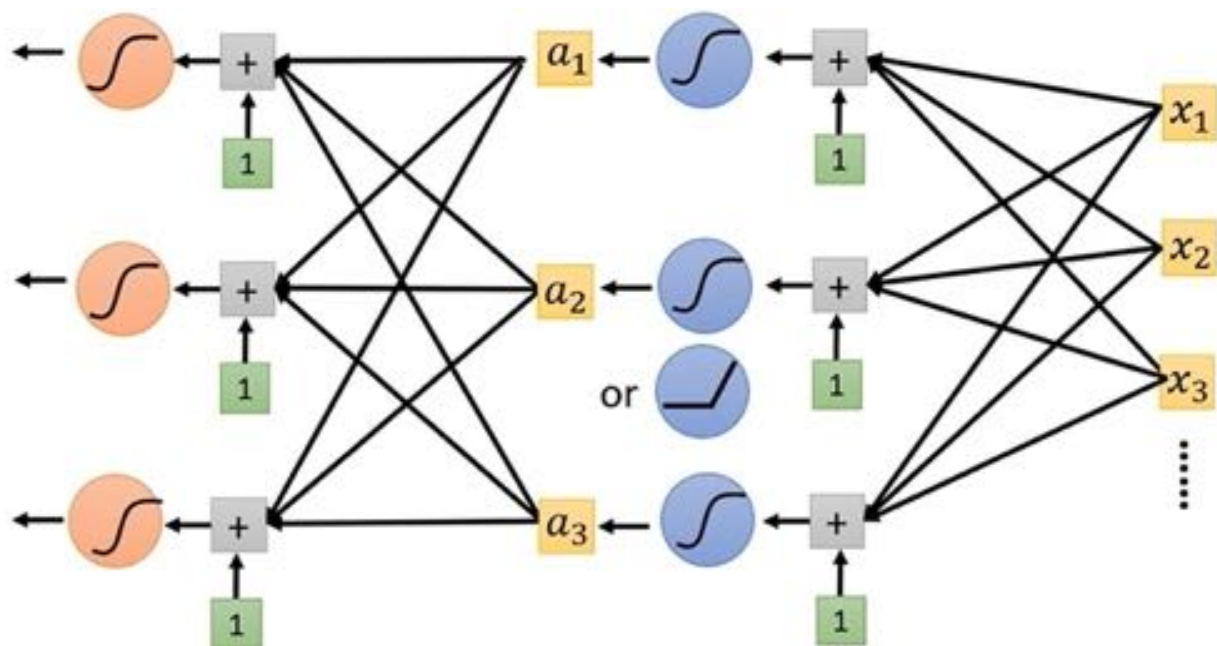
但是，对于720P分辨率，B的MOS分比A要差。

视频的分辨率越低，折损越低越好，所以B比A好吗？



8 视频画质评测还可以做什么？

当我们积累了1年多的评测数据后，我们发现我们还能做更多的事情。得数据者得天下。



1. 不同客观算法的统一测试集

2. 画质增强自动调参系统

3. 基于设备的模型训练

4.

$$a' = \sigma(b' + W'a) \quad a = \sigma(b + Wx)$$

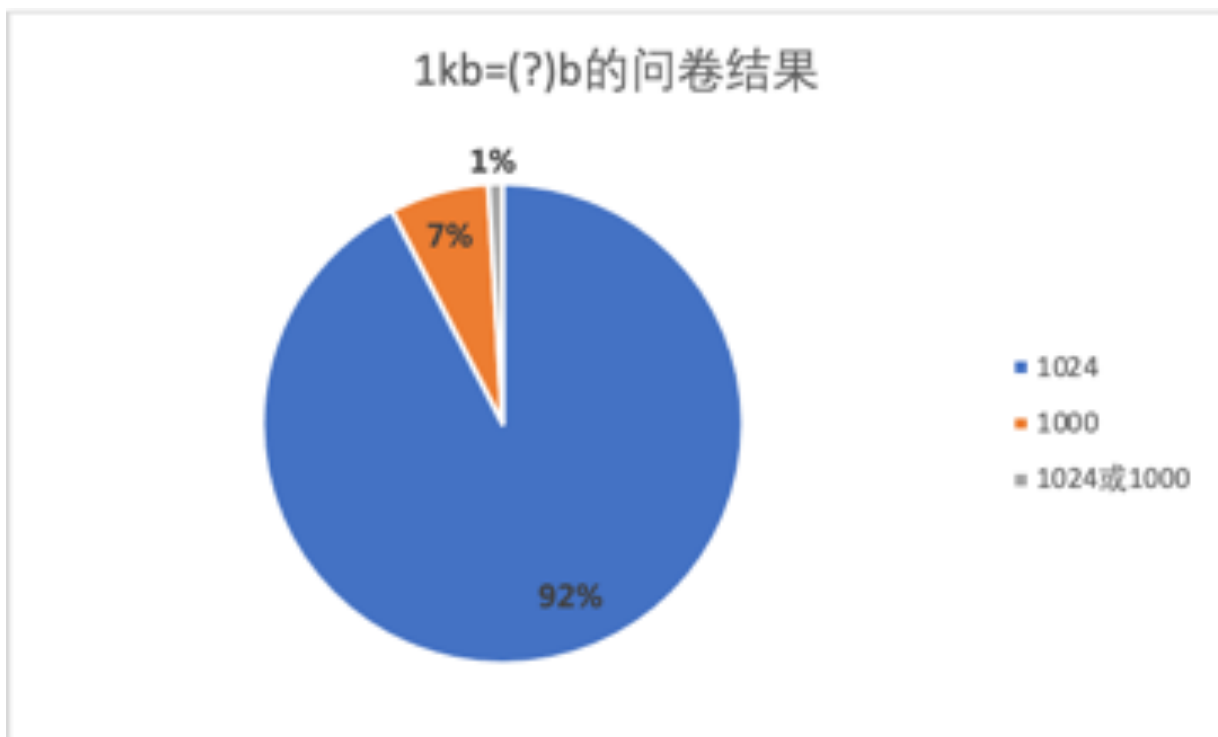
```
ffmpeg_master master ./ffmpeg -i bdd.mp4 -vf visualstats=hsl=1:contrast=1:cont_mode=0:colorfulness=1:texture=1:txtr_mode=0:print_summary=1 -f null -
```

```
[Parsed_visualstats_0 @ 0x7f9f29914340] Frames:125, Calculate Frames:125, H:186.2, S:38.9, L:58.9, contrast:2.7, colorfulness:41.7, texture:6.6
```


9 视频画质评测进展中遇到的有趣的问题

沿途的风景也是那么美丽，也能收获满满~

1. 1kB = 1024B ?



1kB=1000B, 1KiB=1024B

2. 如何计算视频每帧的对比度（图像标准差）？



分块？

全局？

$$\sigma = \sqrt{\frac{1}{N-1} \sum_i (x_i - \bar{x})^2}$$

$$\sigma = \sqrt{\frac{1}{N \times (N-1)} \left(N \sum_i x_i^2 - (\sum x_i)^2 \right)}$$

视频画质评测一路走来的感想——唐僧扫塔

起点低，当下净，回头脏，平常道。



无论是做科研，还是工作，其实都是一样

起点低：把自己放低，扎扎实实从头做起，不要好高骛远

当下净：做好当下的事情，当下没做好谈何过去和未来

回头脏：见识达到一定的高度和广度，回首过去皆是荆棘

平常道：不要总想着做高大上的事，简单的事情做到极致即神话

人总会过高的评价自己，最终是大鱼抓不到，小鱼不想抓，只留下无尽的遗憾和抱怨。



简单事情重复做

重复事情用心做

用心事情持久做



灵镜也已经成长为

公司级别的平台~

THE ART OF
**PIRATES OF THE
CARIBBEAN**
ON STRANGER TIDES
IN ASSOCIATION WITH
Disney

Disney
PIRATES OF THE CARIBBEAN
ON STRANGER TIDES
IN THEATERS 5.20.11

扬帆

起航！！！！

BLACKBEARD'S SHIP,
THE QUEEN ANNE'S REVENGE