

# WebRTC Con

为开发者赋能  
为行业加速

2018. 5. 19 - 20 · 上海光大会展中心

更多分享 敬请期待 ...

邱建林 - 英特尔实时通信客户端架构师  
《WebRTC 之 H.264/265 硬件编解码优化》

刁磊 - 哒哒英语研发工程师  
《WebRTC 在直播场景问题和优化》

陈域将 - 阿里巴巴高级技术专家  
《WebRTC 在阿里巴巴的实践》

张轲 - 腾讯音视频实验室高级工程师  
《腾讯语音通信 QoS 优化实践》

冯巍 - 爱奇艺技术产品中心研究员  
《爱奇艺 WebRTC 在智能音箱领域的应用》

吴涛 - 陌陌视频流媒体技术负责人  
《低延迟音视频传输技术在直播领域的应用》

和君 - TutorABC 大前端负责人

吴威麒 - 触宝电话音频技术专家  
《触宝电话如何改善通话质量》

姜健 - Google 软件工程师

来一次纯正的 WebRTC 之旅



即刻享受 8 折优惠

LiveVideoStack  
— 音视频技术社区 —

# WebRTC Con

为开发者赋能  
为行业加速

2018. 5. 19 - 20 · 上海光大会展中心

更多分享 敬请期待 ...

陈功 - 声网首席 WebRTC 架构师

《声网的 WebRTC 服务架构与实践》

周正宁 - Aupera 傲睿智存 CTO

章琦 - 唐桥科技首席架构师

《基于 WebRTC 的多方视频会议融合架构设计》

吴涛 - 陌陌视频流媒体技术负责人

冯巍 - 爱奇艺技术产品中心研究员

《爱奇艺 WebRTC 在智能音箱领域的应用》

Zoe Liu - Google 软件工程师

《AV1 + WebRTC 助力开发者》

盛骁杰 - 优酷 VR 技术专家

朱浩齐 - 网易云易盾 CTO

Niklas Blum - Google PM of WebRTC

《Google 解读 WebRTC 优化与改进》

来一次纯正的 WebRTC 之旅



即刻享受 9 折优惠

LiveVideoStack  
— 音视频技术社区 —

# LiveVideoStack Meet 多媒体开发新趋势

鲍金龙

2018年3月31日

微软大厦 东直门会议室



# 目录

1

超分辨率背景

2

几年来的进展

3

R  
A  
I  
S  
R

4

B  
S  
R

5

V  
M  
A  
F

# 知识点

● ..... ● 视频的超分辨率

● ..... ● 机器学习与神经网络

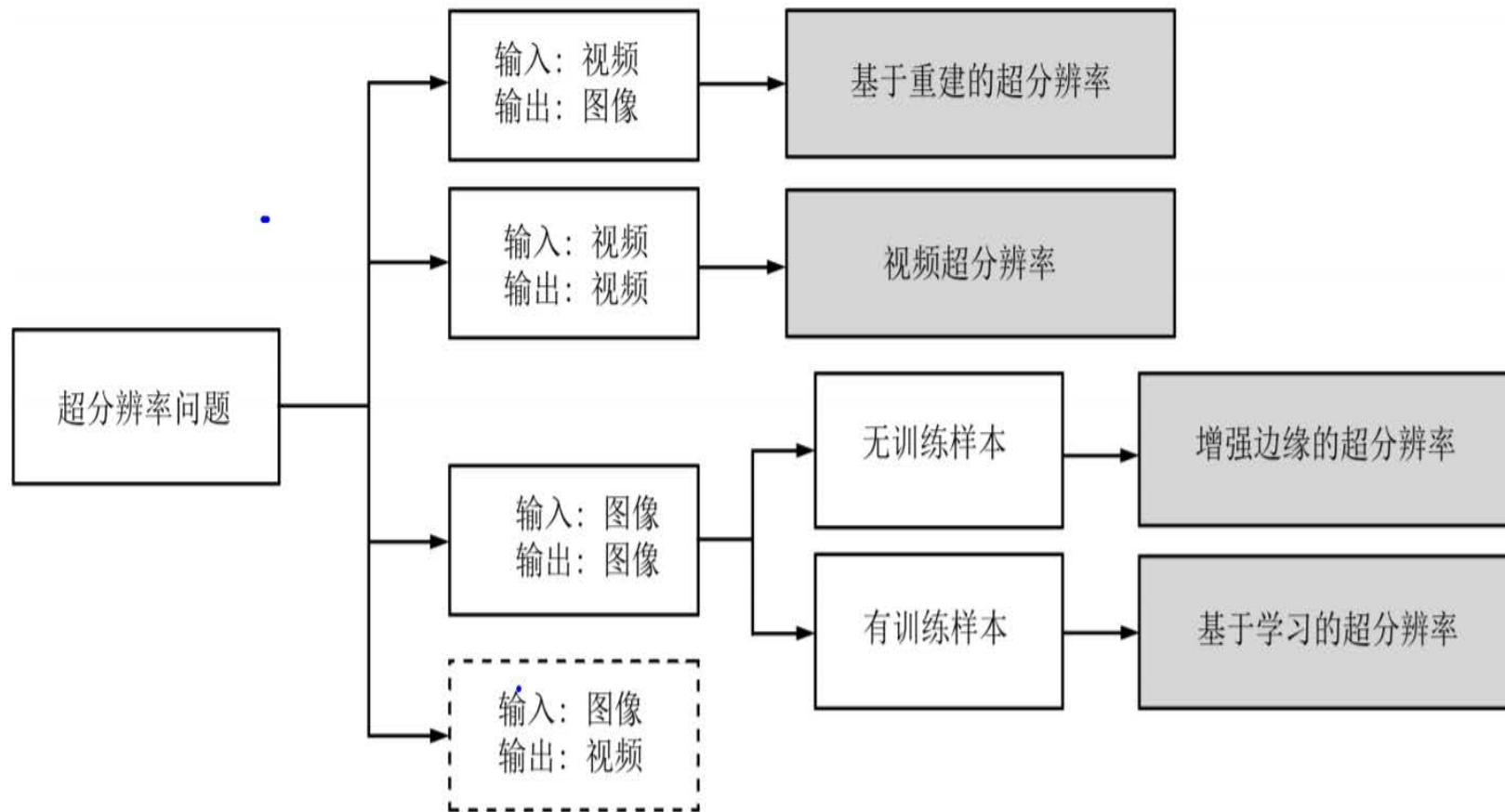
3月31日  
微软大厦

● ..... ● 图象模式的提取

● ..... ● 人类视觉心里模型

● ..... ● VMAF

# 超分辨率的分类



# 现代绘画流派：超写真



# XDOG



(a) Source



(b) Sobel



(c) Canny



(d) LoG/DoG zero-crossings



(e) Thresholded DoG



(f) XDOG

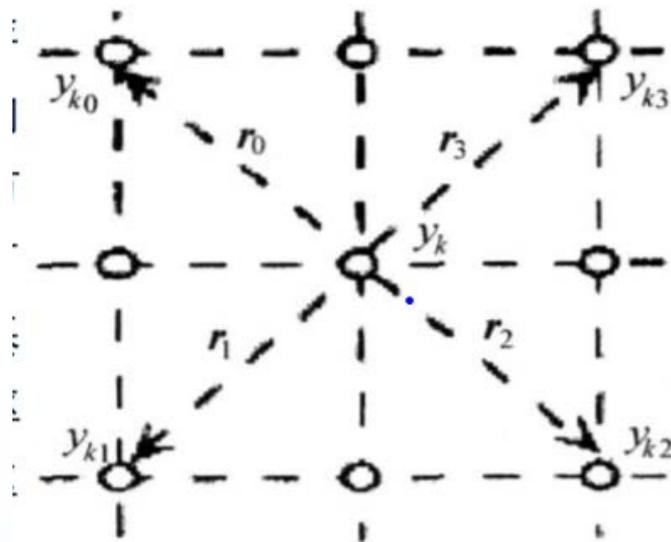


(g) Flow-XDoG

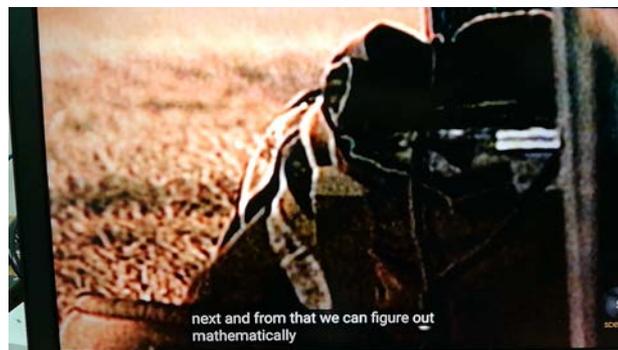
# 单帧图像的超分辨率

高低分辨率图像间存在以下关系:

- 1) 边缘取向具有分辨率不变性;
- 2) 协方差矩阵 $R$ 和协方差矢量含有足够多的边缘取向信息。基于这两点,可以用低分辨率协方差的特性指导高分辨率插值。



# Video: NASA未公开的秘密



## CENTENNIAL OLYMPIC PARK

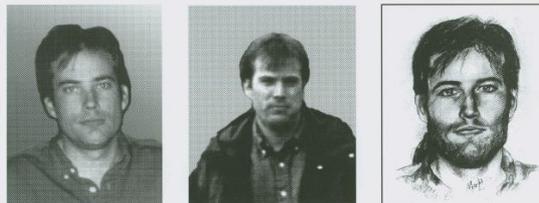
July 27, 1996



## FBI TEN MOST WANTED FUGITIVE

MALICIOUSLY DAMAGED, BY MEANS OF AN EXPLOSIVE DEVICE, BUILDINGS AND PROPERTY AFFECTING INTERSTATE COMMERCE WHICH RESULTED IN DEATH AND INJURY

### ERIC ROBERT RUDOLPH



Date of photograph October 1997      Date of photograph October 1997      Artist rendition July 1998  
Aliases: Bob Randolph, Robert Randolph, Bob Rudolph, Eric Rudolph and Eric R. Rudolph.



# 视频超分辨率

- 1) 视频的超分辨率重建是指利用低分辨率图像帧之间的相似性、冗余性以及一些先验知识进行数据融合, 以得到高分辨率的序列图像.
- 2) 超分辨率重建一般包含两个步骤: 首先将各个低分辨率图像进行配准, 即运动估计过程; 然后利用频域或空域重建算法将已配准的低分辨率图像融合为一幅高分辨率图像.
- 3) 视频超分辨率重建中, 配准精度的高低对重建效果的影响是巨大的, 一个不好的配准结果往往会 导致重建的失败.
- 4) 运动估计的方法通常可分为三类: 空域、频域和空频域方法. 超分辨率重建中采用最多的是空域法中基于光流的方法和基于块的方法.

# 频域增强方法的过程

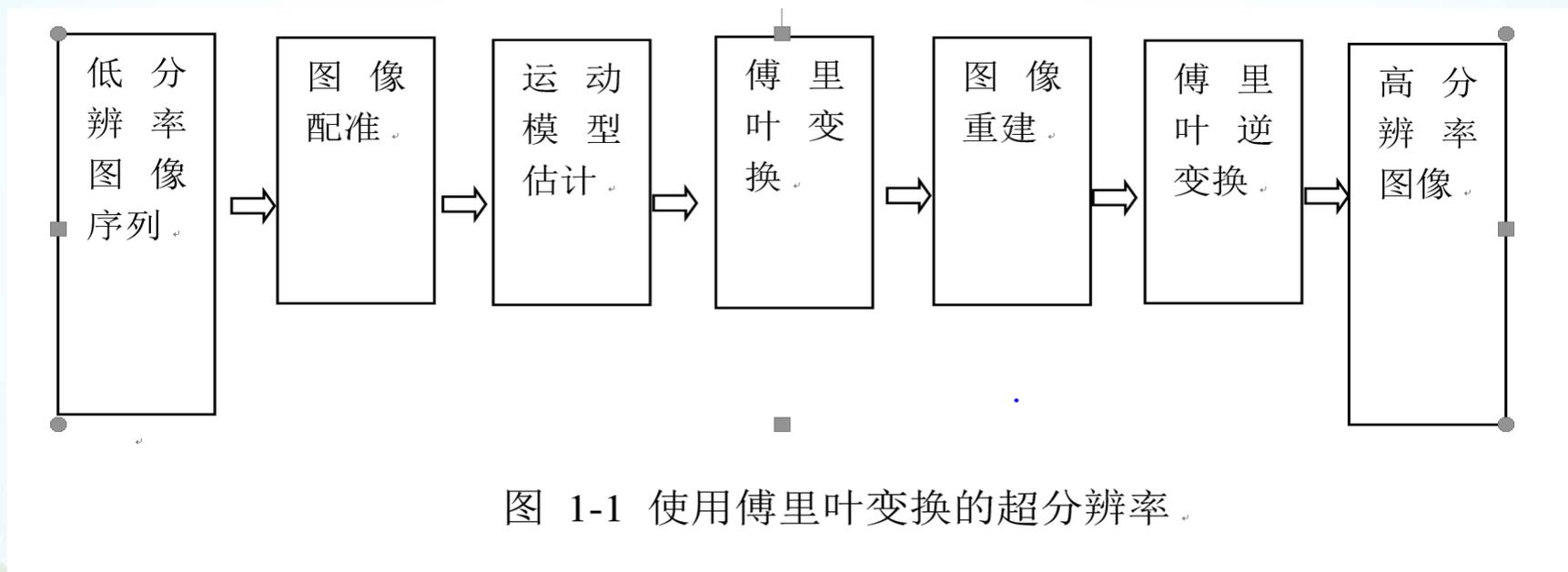
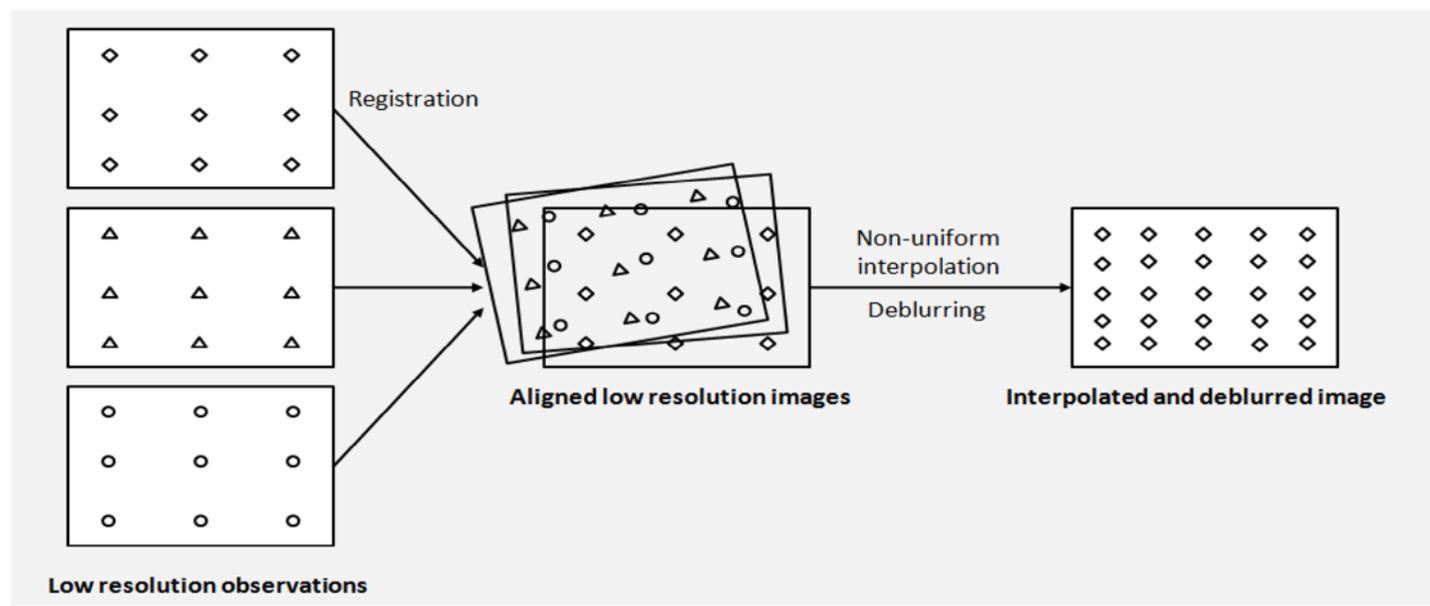


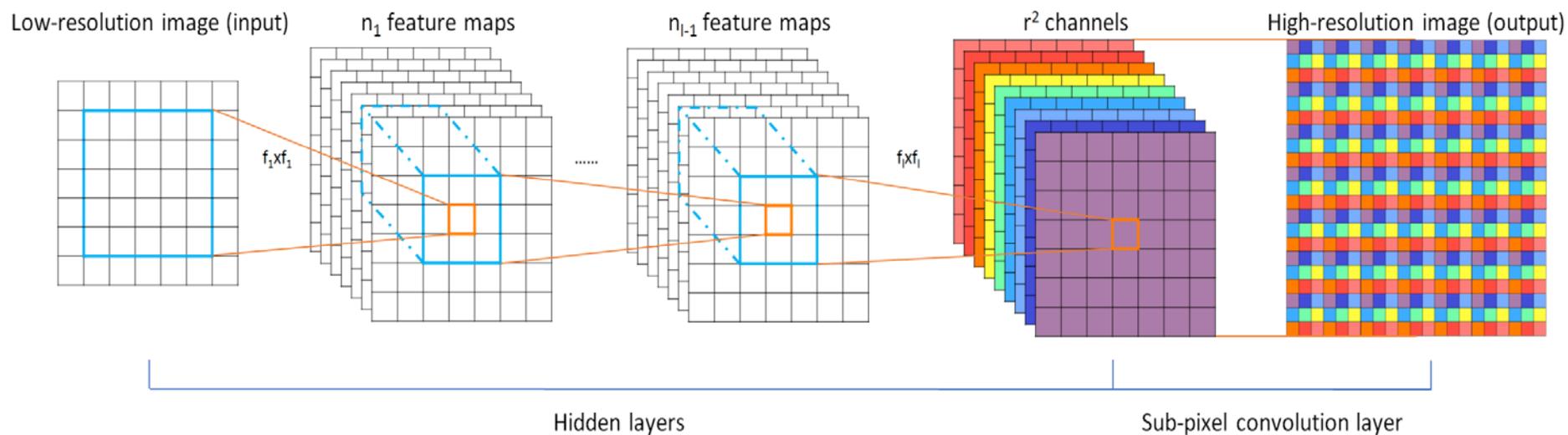
图 1-1 使用傅里叶变换的超分辨率。

# 图像配准

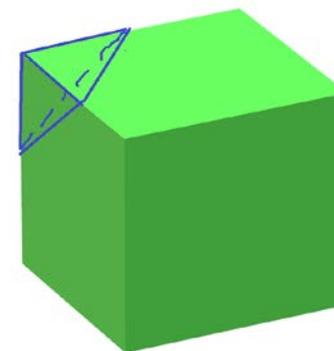
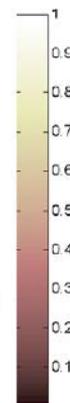
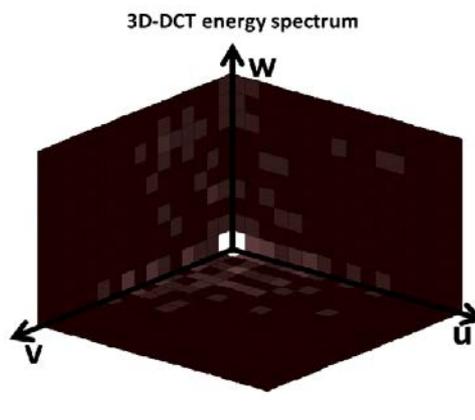
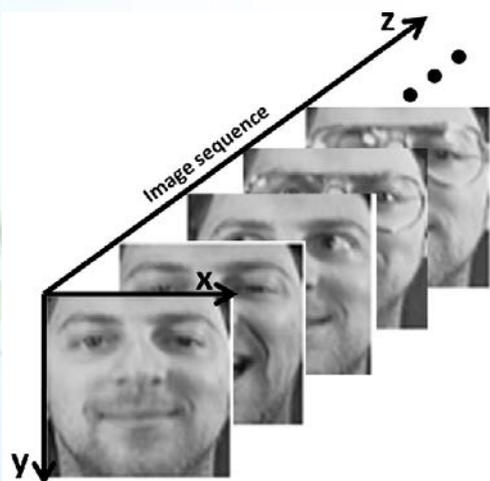
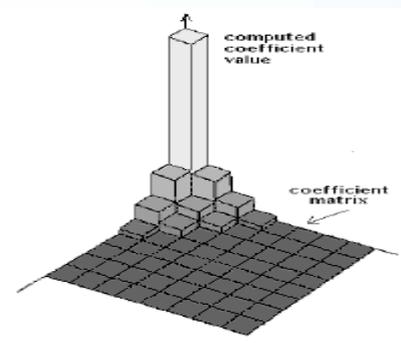
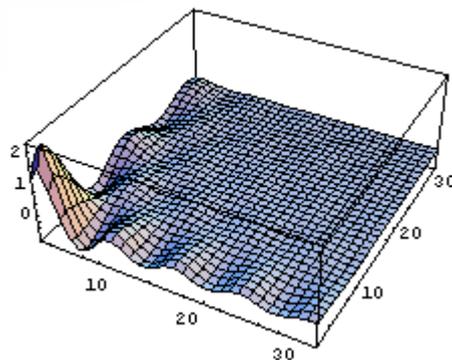
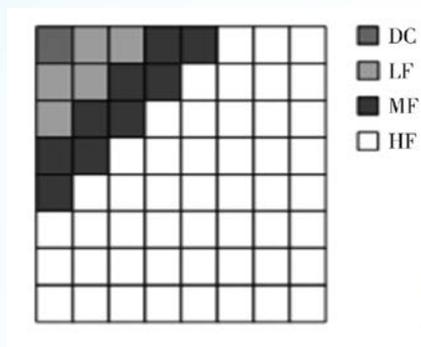
超分辨率重建一般包含两个步骤: 首先将各个低分辨率图像进行配准, 即运动估计过程; 然后利用频域或空域重建算法将已配准的低分辨率图像融合为一幅高分辨率图像. 视频超分辨率重建中, 配准精度的高低对重建效果的影响是巨大的, 一个不好的配准结果往往会 导致重建的失败



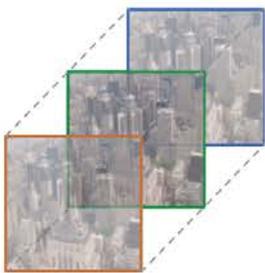
# 子像素提取



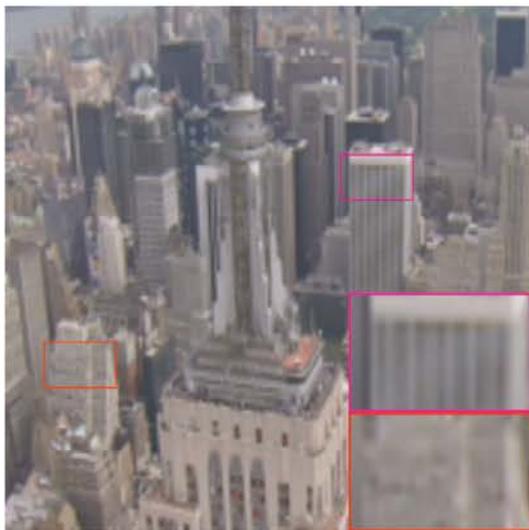
# 变换方法的能量分布



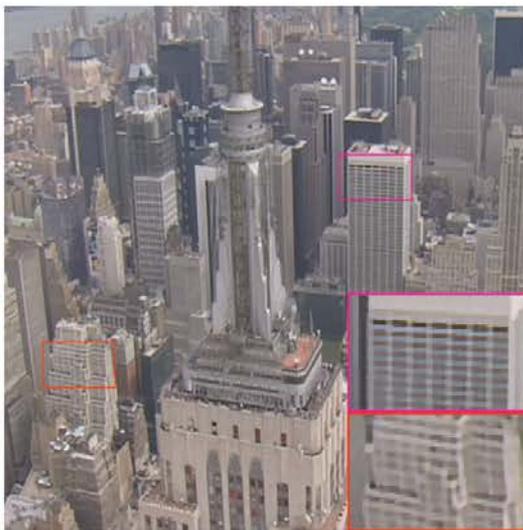
# Bayesian 过程



(a) Input low-res



(b) Bicubic up-sampling  $\times 4$

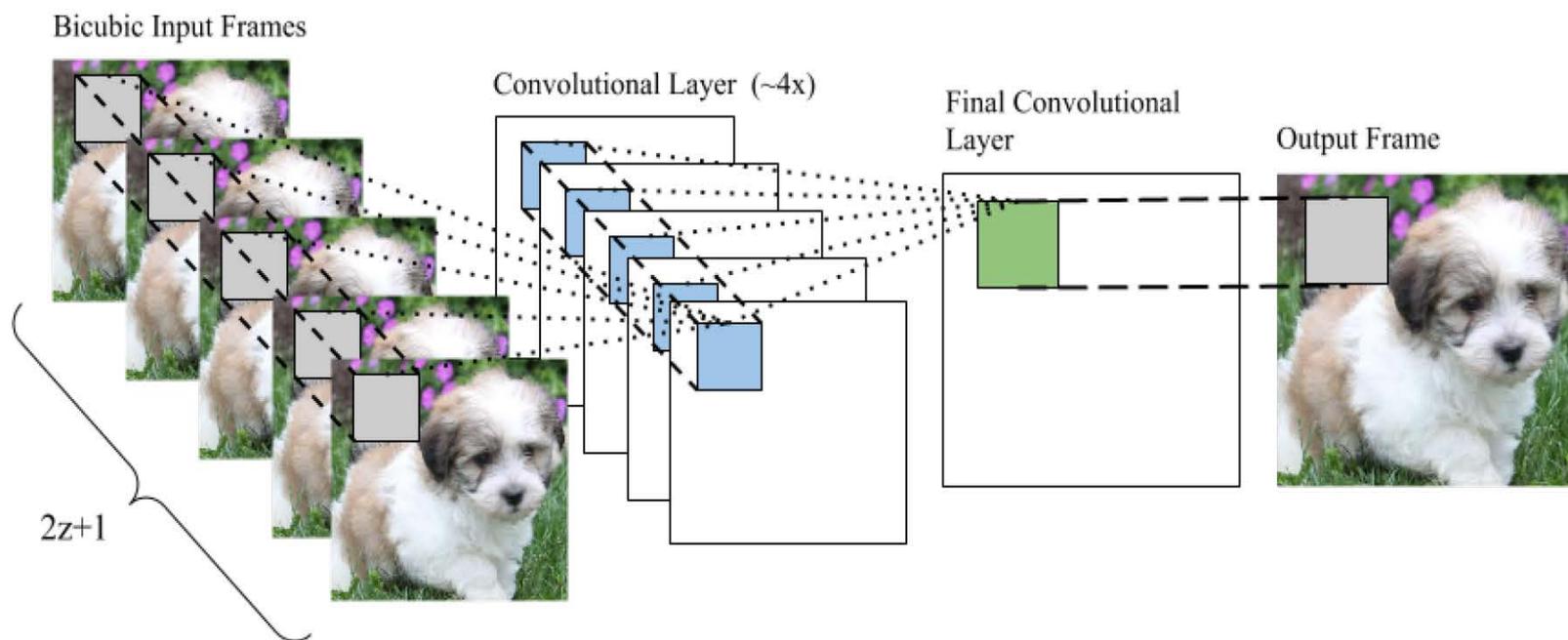


(c) Output from our system

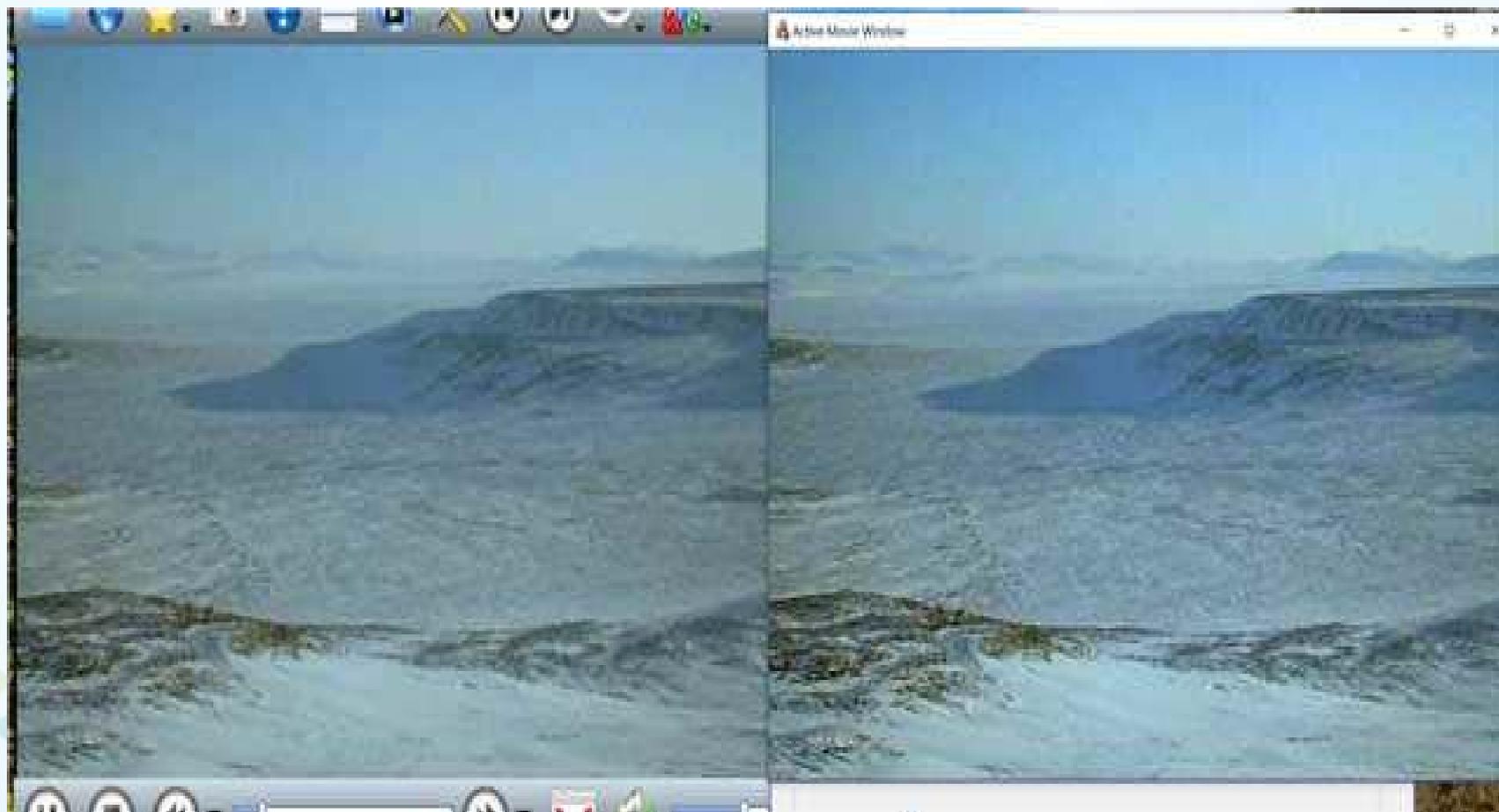


(d) Original frame

# 使用CNN神经网络的例子



# 高斯降噪的表现 720x400



# 高斯降噪的表现 1280x720

The screenshot shows a video player window titled "m.Series.4.Viking.Secrets.720p.HDTV.x264.AAC.mp4[eztv].mp4 - SMPlayer". The video content displays a news article with the headline "VIKINGS REVEALED: New Discoveries to Rewrite Viking History". The article text includes: "Top Danish archaeologists are working overtime to keep up with the incredible discoveries being made at not one but two of the most exciting new dig sites to be uncovered in a generation or more. The discovery of an entire ring fortress in Koge and the excavation of the first Viking marketplace are both changing everything historians thought they knew about the legendary Scandinavian rulers known the world over as the Vikings. In 2014, LARSEN's team confirmed what some elite archaeologists had suspected as an yet undiscovered ring fortress built under the direction of King Harald Bluetooth. Located on the outskirts of small town Koge, the fortress is an unexpected as it is supposed. It has all just 3 miles northwest of Thorsmark. It is considered a major find, as it shows that the Vikings were not just a nomadic people but also a settled society. The discovery of the fortress is a major find of the Viking era. It shows that the Vikings were not just a nomadic people but also a settled society. The discovery of the fortress is a major find of the Viking era. It shows that the Vikings were not just a nomadic people but also a settled society." A Gaussian noise effect is applied to the video, making the text difficult to read. A white text overlay at the bottom of the video frame reads "Archeologists are making major discoveries". The video player interface includes a progress bar showing 00:01:53 / 00:42:03. The Windows taskbar at the bottom shows the system tray with the date 2018/3/19 and time 17:47.

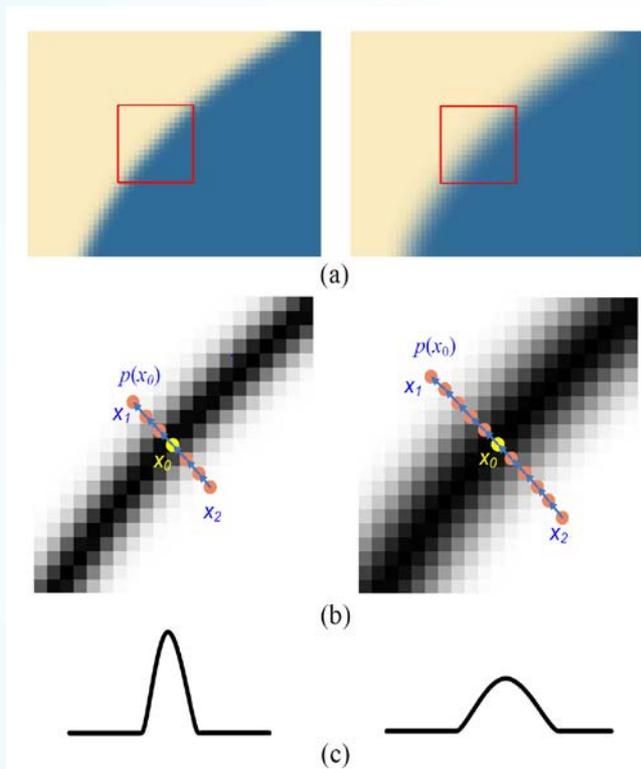
# 图像超分近来的发展

- 1) 无论频域还是空域滤波器，都越来越局部化，跟图像的模式相关。
- 2) 图像模式的获取分为稀疏矩阵法和变换法两种。
- 3) 滤波器由少量预设定类型进化为机器学习训练类型。
- 4) 算法速度越来越快，预计很快会出现在移动端实时运行的SR.

# 机器学习与深度卷积神经网络

- 1) 卷积是介于空域插值方法和频域变换滤波方法之间的一种形态，兼具两者的特征。
- 2) 多层的卷积神经网络可以模拟比较复杂的非线性函数。
- 3) DNN, CNN, RNN.

# 径向梯度变换和A+, SRCNN



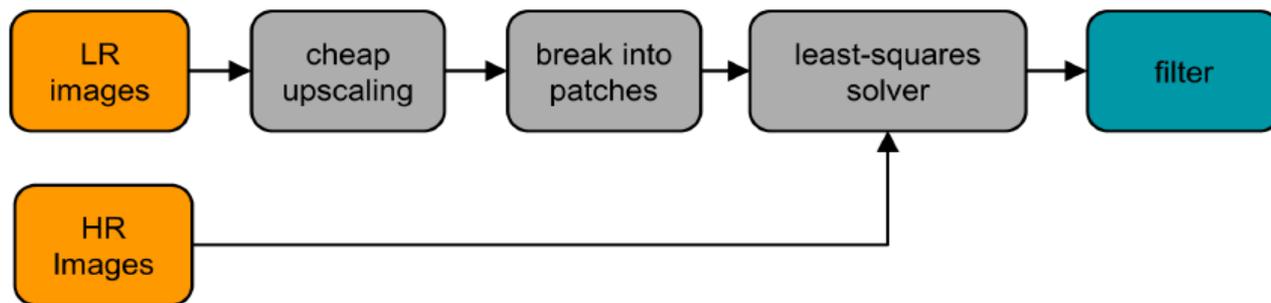
A+:

- 1) 根据周围像素生成模式词典;
- 2) 使用稀疏矩阵压缩方法生成HASH;
- 3) 根据HASH键值选择滤波器。

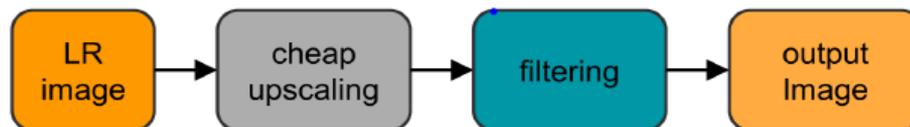
SRCNN: 使用稀疏矩阵的HASH来选择优质滤波器的方法, 可以认为是一个卷积神经网络的工作过程。

# RAISR

传统的全局性的滤波器表现不够出色。  
根据局部内容的特性选择滤波器成为主流趋势。  
RAISR是近年来比较突出的成果。

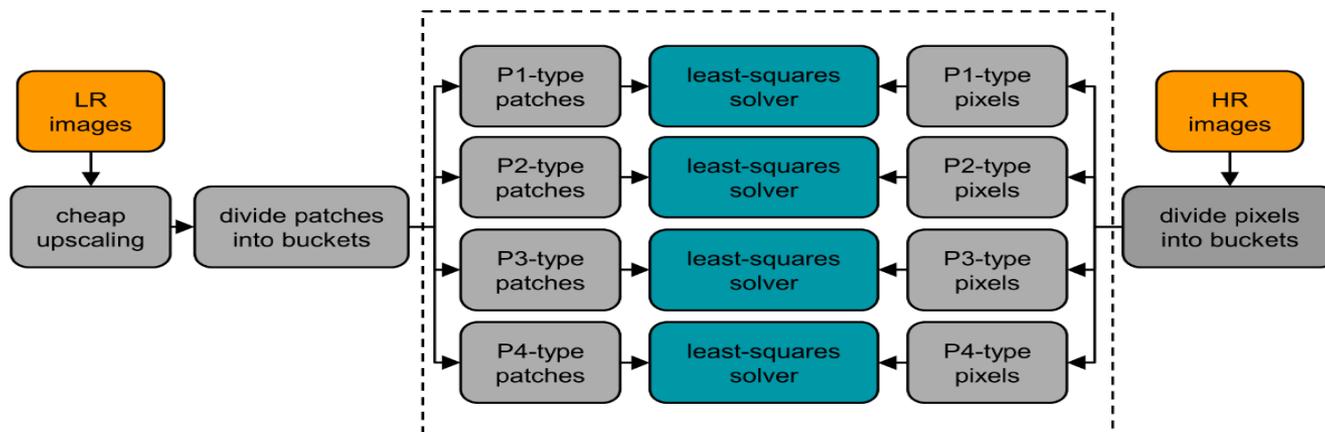


(a) Learning Stage

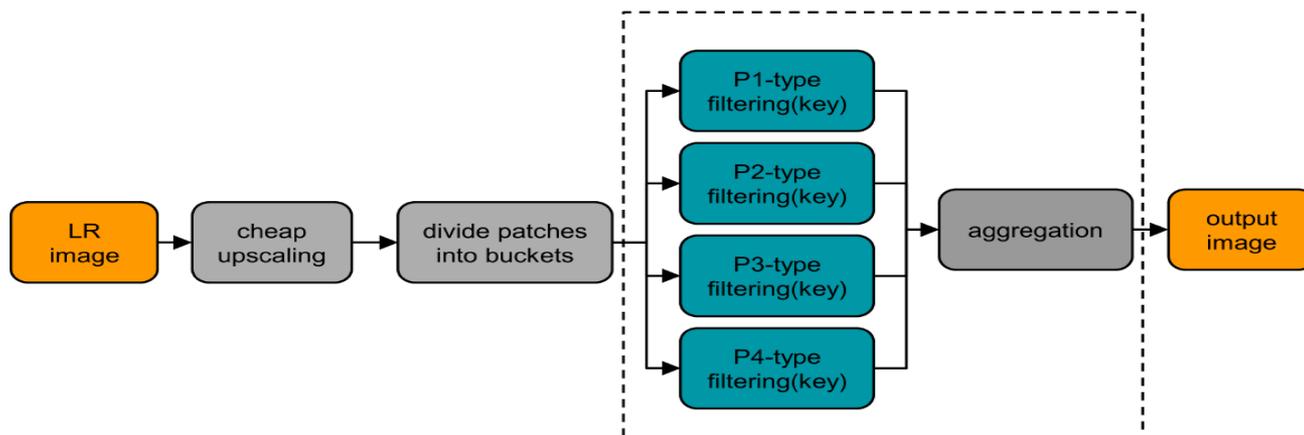


(b) Upscaling Stage

# BUCKETS

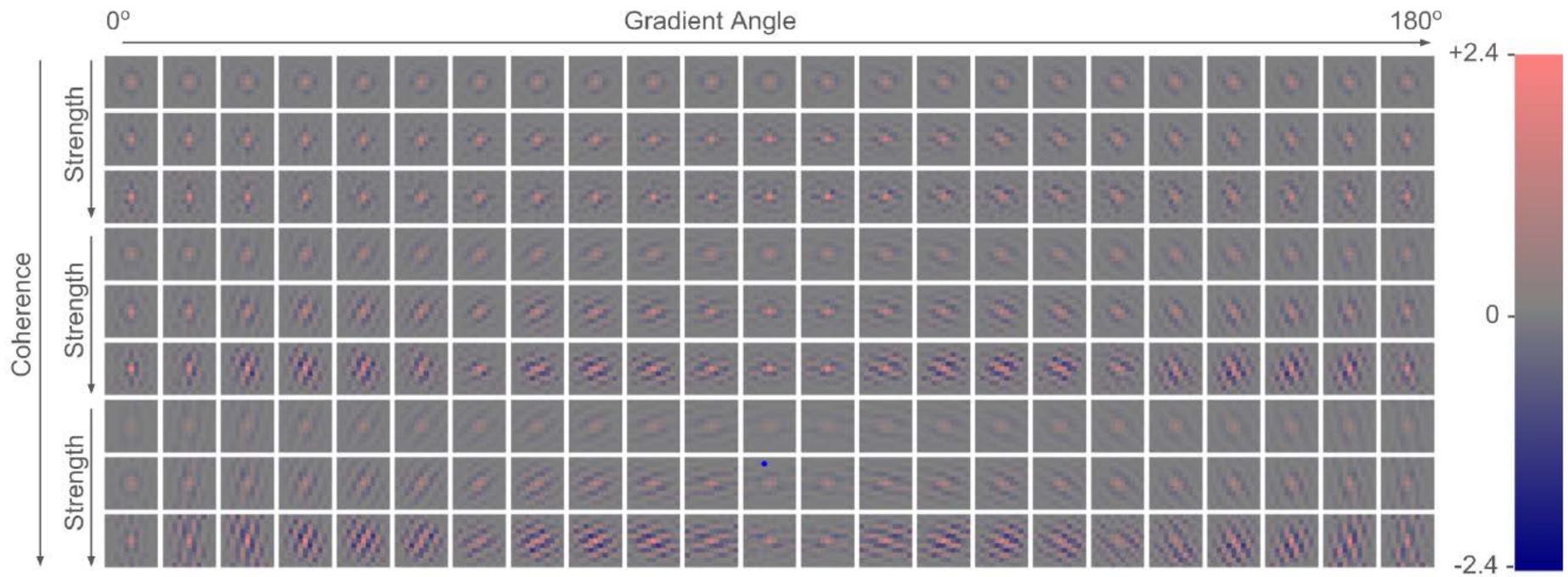


(a) Learning Stage



(b) Upscaling Stage

# RAISR的filter



(a) 2× upscaling filters

## RAISR的CT

$a$	$d$	$g$
$b$	$e$	$h$
$c$	$f$	$i$

(a)

$e > a$	$e > d$	$e > g$
$e > b$	-	$e > h$
$e > c$	$e > f$	$e > i$

(b)

100	80	90
105	95	85
110	120	170

(c)

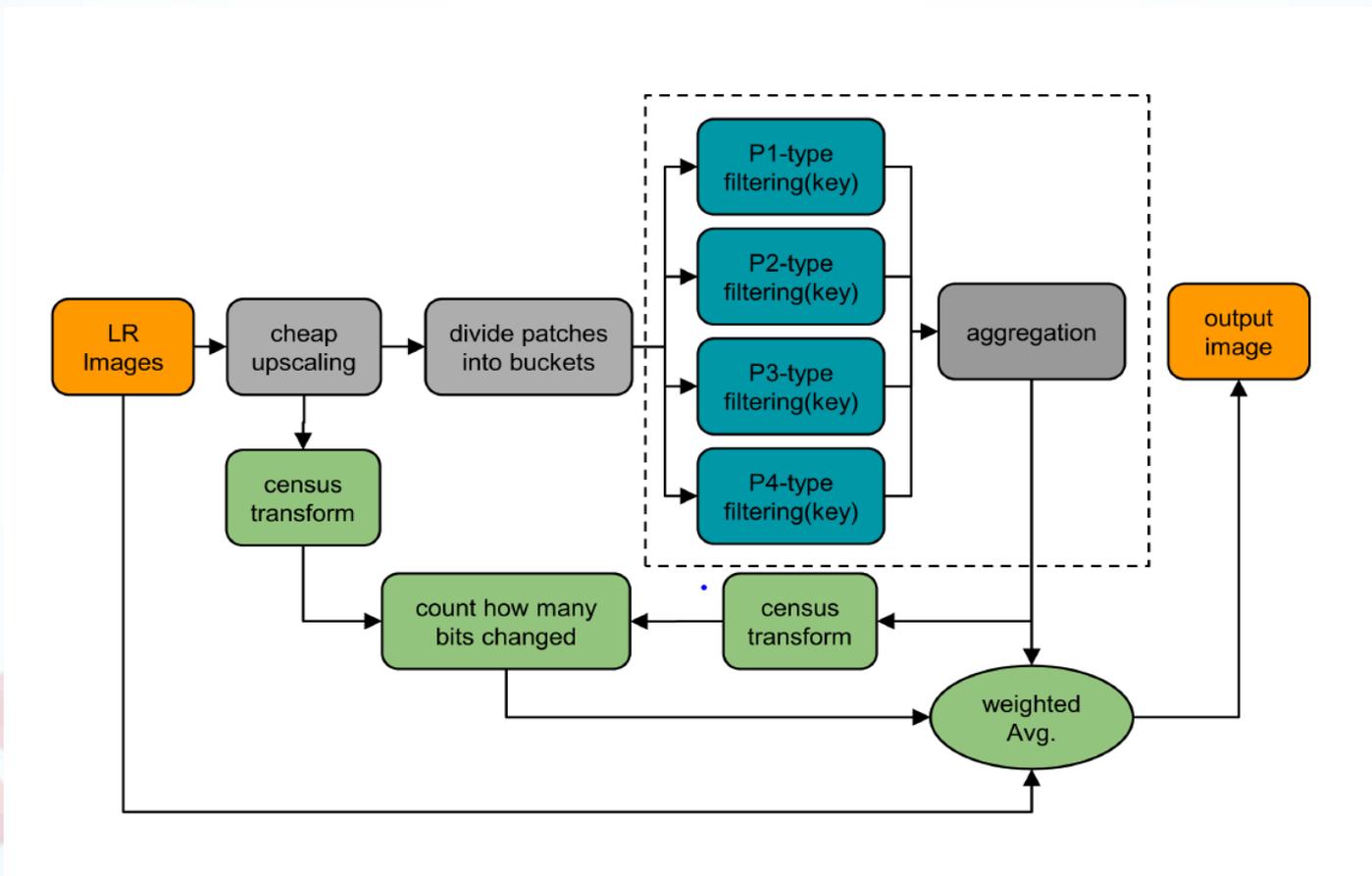
0	1	1
0	-	1
0	0	0

(d)

8 bit string							
$a$	$b$	$c$	$d$	$f$	$g$	$h$	$i$
0	0	0	1	0	1	1	0

(e)

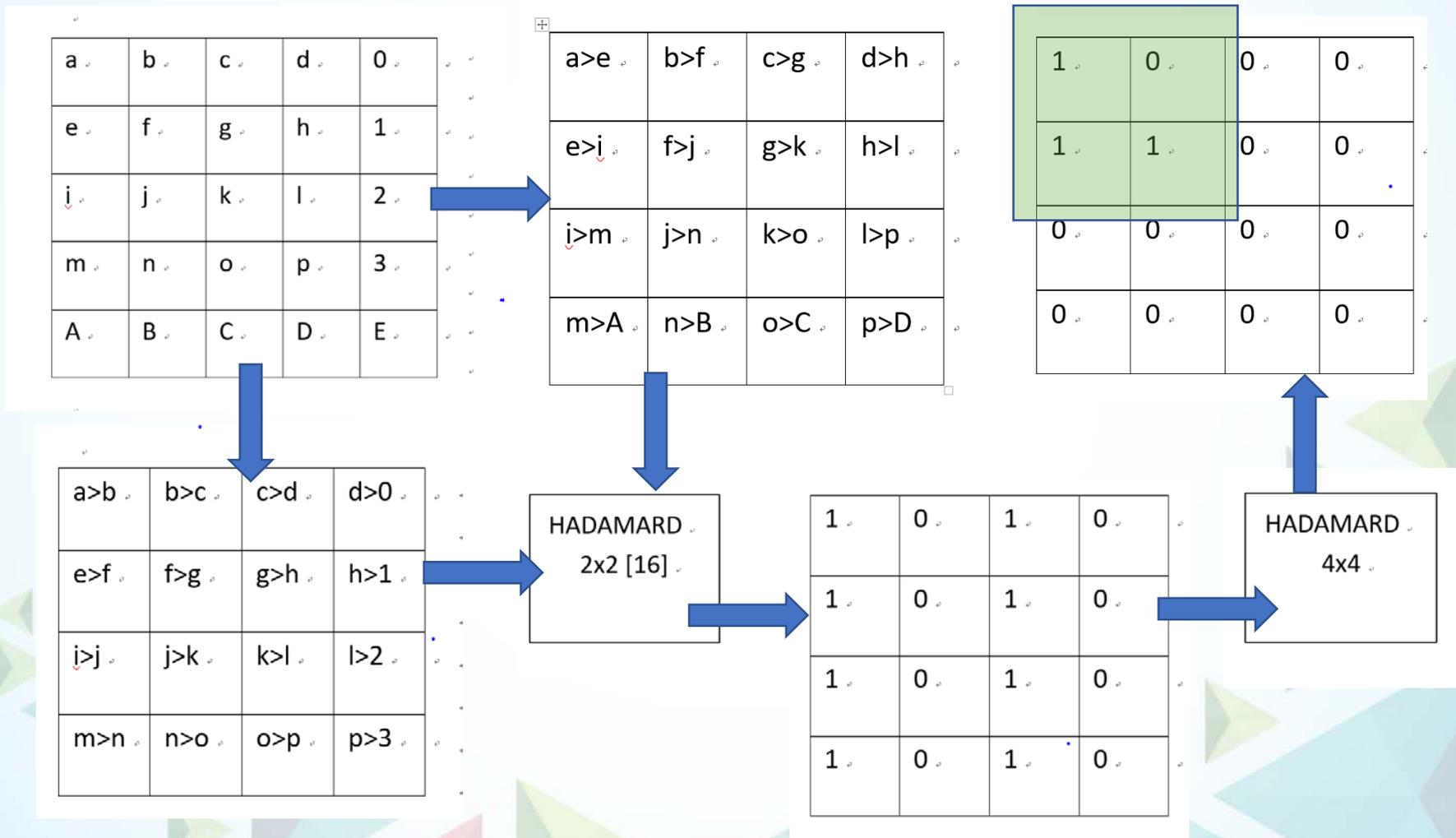
# CT的作用



# BSR的需求

- 1) 3x3的高斯降噪，在720P以上分辨率中体现的比较理想。  
低分辨率图像表现不佳。
- 2) 基本的BICUBIC拉伸算法质量不能满足需求。
- 3) 寻找一种快速有效的拉伸算法，对图像的梯度有更好的保持。
- 4) 由此产生了思路：对像素微环境模式进行匹配，产生带有局部方向性的双三次曲线滤波器。
- 5) 经过有效拉伸后再进行高斯滤波。
- 6) 跟RAISR的CT过程类似；但是PATTERN的识别和编码更加复杂。滤波器相对简单许多。

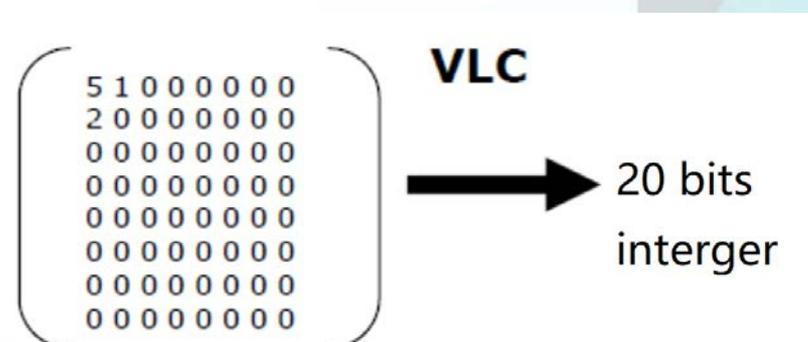
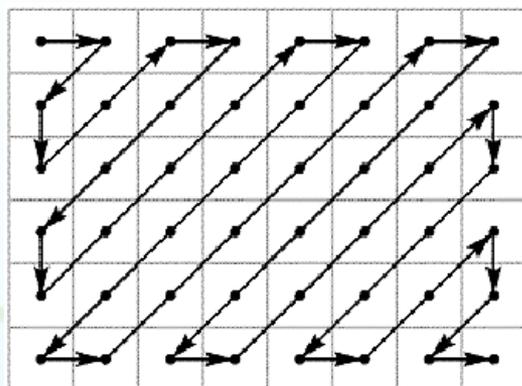
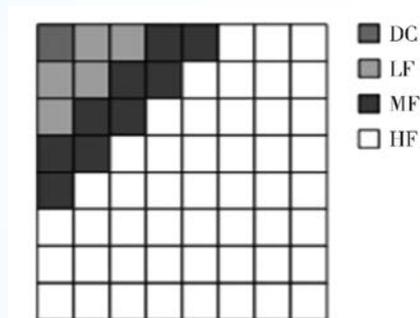
# 像素所属环境的微小模式



# BSR 的模式提取

$$H_4 = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & -1 & 1 & -1 \\ 1 & 1 & -1 & -1 \\ 1 & -1 & -1 & 1 \end{bmatrix}$$

$$H_8 = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & -1 & 1 & -1 & 1 & -1 & 1 & -1 \\ 1 & 1 & -1 & -1 & 1 & 1 & -1 & -1 \\ 1 & -1 & -1 & 1 & 1 & -1 & -1 & 1 \\ 1 & 1 & 1 & 1 & -1 & -1 & -1 & -1 \\ 1 & -1 & 1 & -1 & -1 & 1 & -1 & 1 \\ 1 & 1 & -1 & -1 & -1 & -1 & 1 & 1 \\ 1 & -1 & -1 & 1 & -1 & 1 & 1 & -1 \end{bmatrix}$$



# BSR的滤波器设计思路

1	-5	-5	1
-5	25	25	-5
-5	25	25	-5
1	-5	-5	1

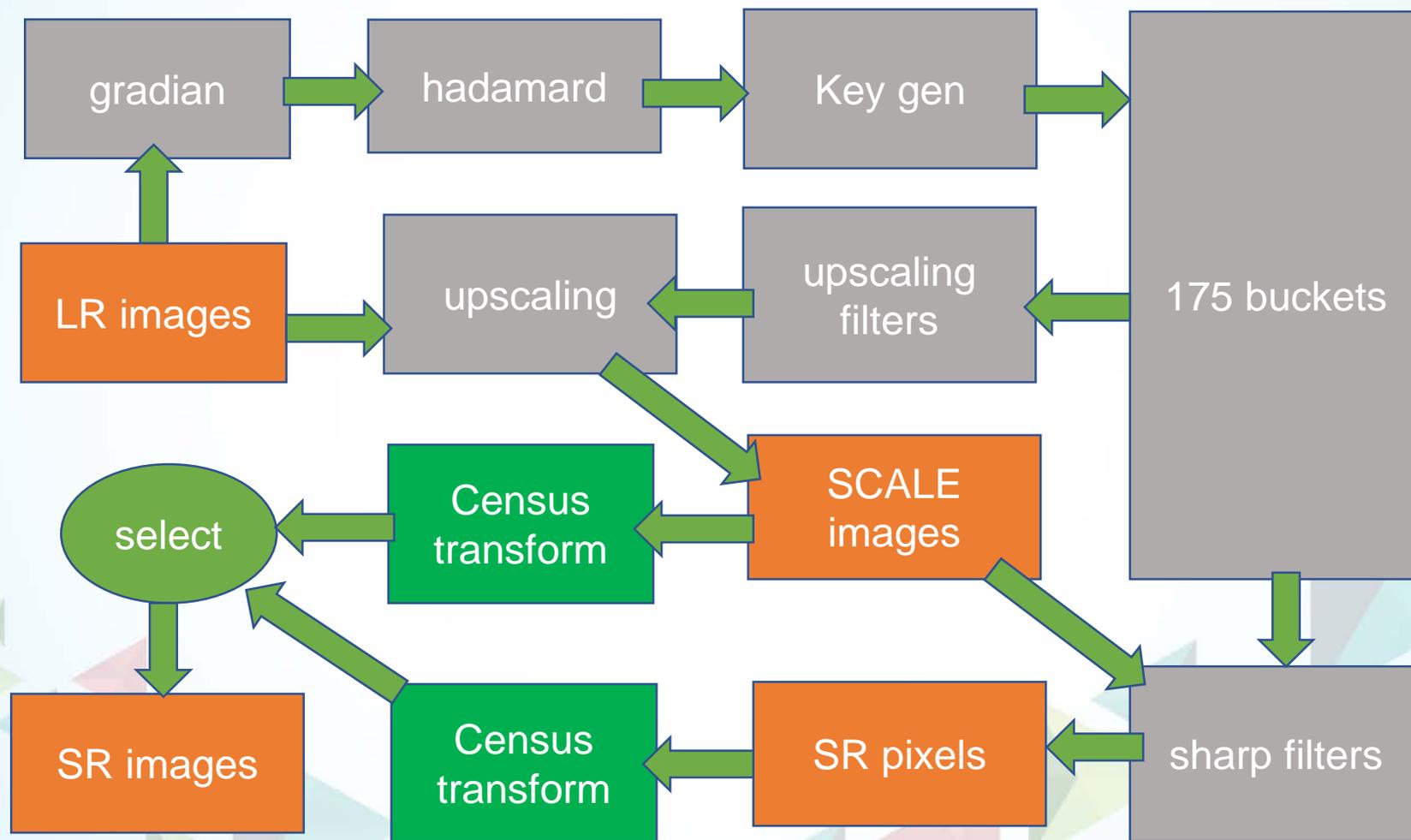
<b>P00</b>	<b>P01</b>	<b>P02</b>	<b>P03</b>
P10	P11	P12	P13
P20	P21	P22	P23
P30	P31	P32	P33

0	0	0	-8
0	0	40	0
0	40	0	0
-8	0	0	0

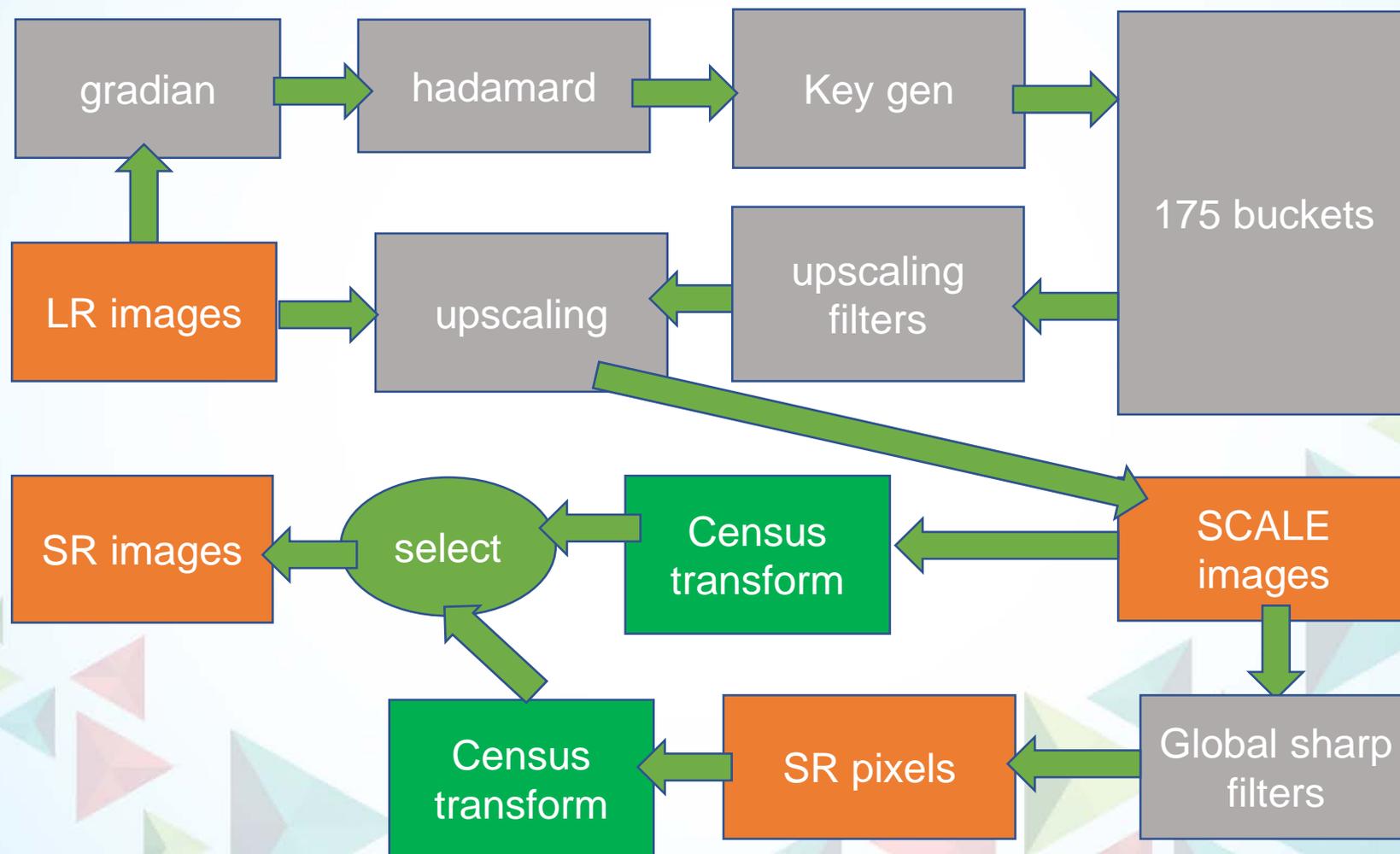
0	0	0	0
-8	40	0	0
0	0	40	-8
0	0	0	0



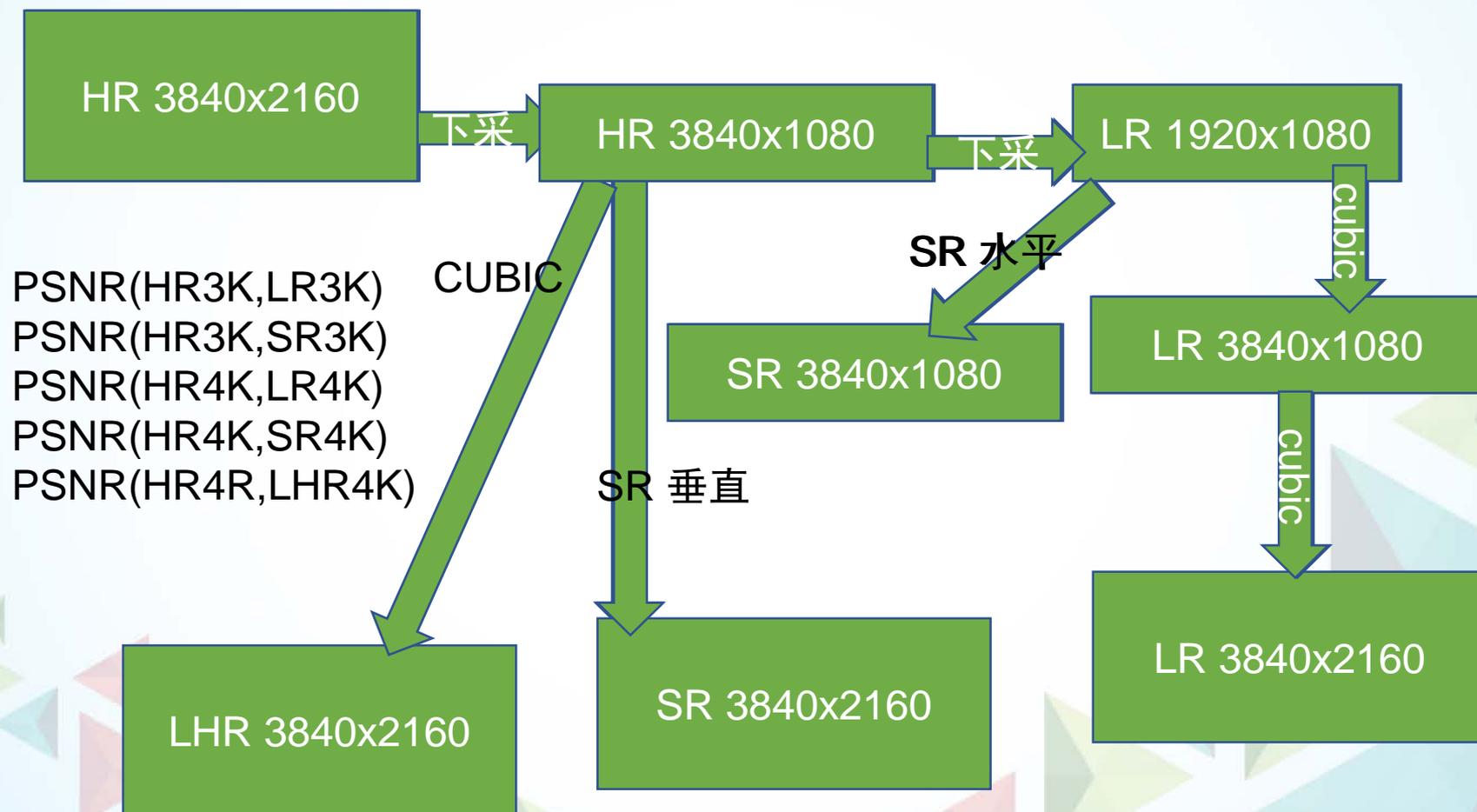
# BSR的拉伸



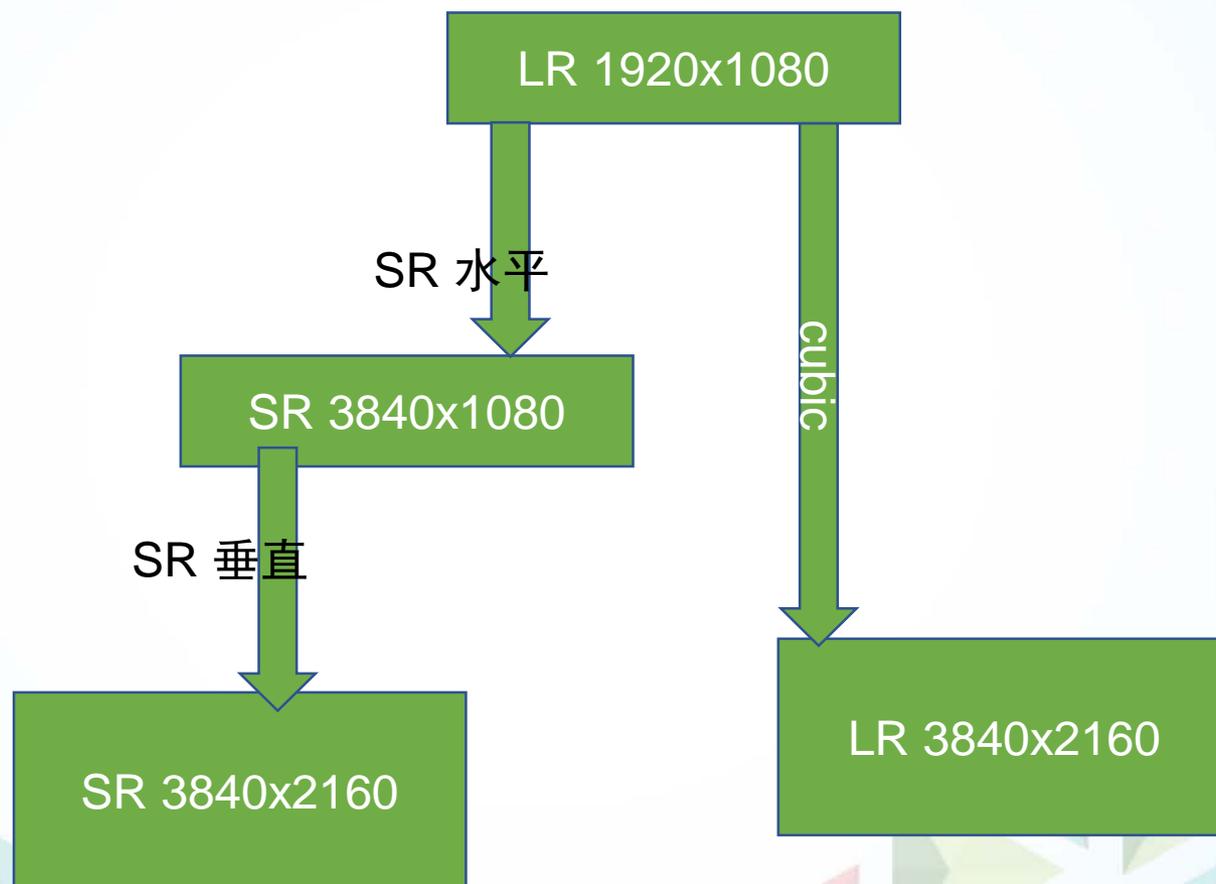
# 简化版BSR.C



# BSR的训练方法

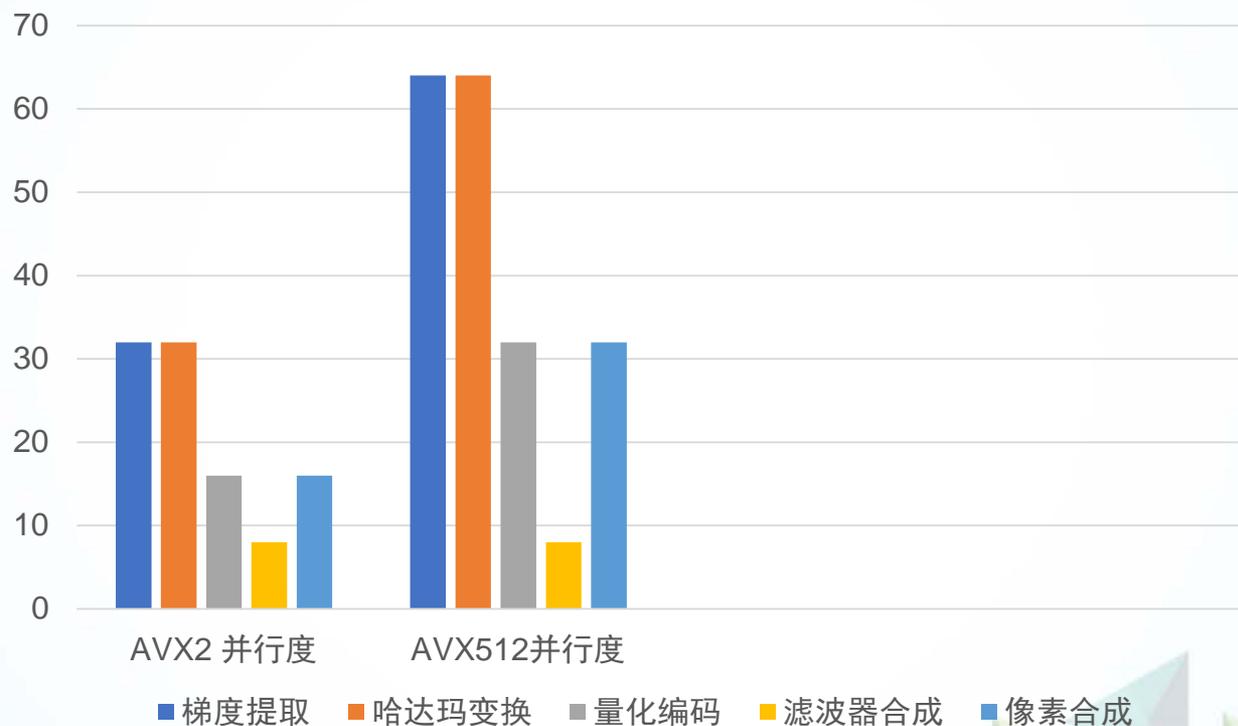


# BSR的推理



# BSR 的SIMD优化

## SIMD指令并行度



# BSR AVX2 训练

4K	FPS 1 thread	FPS 32 threads	PSNR
FAST CUBIC	>200	NA	22.56
BSR	0.25/1	8/32	23.62

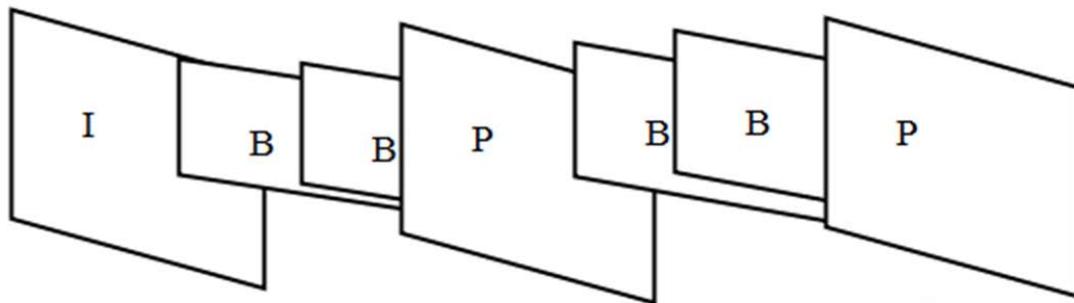
# BSR AVX2 拉伸

1080P->4K	FPS 1 thread	FPS 4 threads
FAST CUBIC	>200	NA
BSR	20	80

# BSR的改进

- 1) 目前实现了简化版的BSR.
- 2) 下一步把降噪过程也跟pattern绑定在一起，做到方向化，模式化。AMSDOG可能是理想的滤波器选择。
- 3) 在GPU加速和NEON/HVX, NPU上做更多的优化工作。

# 人类视觉心理模型



非关键帧进行半幅解码，输出一个垂直方向上分辨率减半的图像。在显示过程中，对减半图像进行BICUBIC拉伸复原到原始分辨率。经过大量测试，结果表明人眼对这种播放序列不敏感，用户很难将原始分辨率标准解码和优化解码区分开来。

# 人类智能对模式的识别和存储

- 1) 人类的大脑特别擅长视觉计算。
- 2) 尤其擅长模式的提取，识别和存储。
- 3) 人类的大脑是一个分布式的，计算和存储一体化的生物计算机。
- 4) 目前可以合理猜测，模式的存储基本上和物体的分辨率无关的。
- 5) 模式包括线条，纹理，运动。影响模式的因素包括噪声，亮度，对比度，复杂度，运动，心理因素。

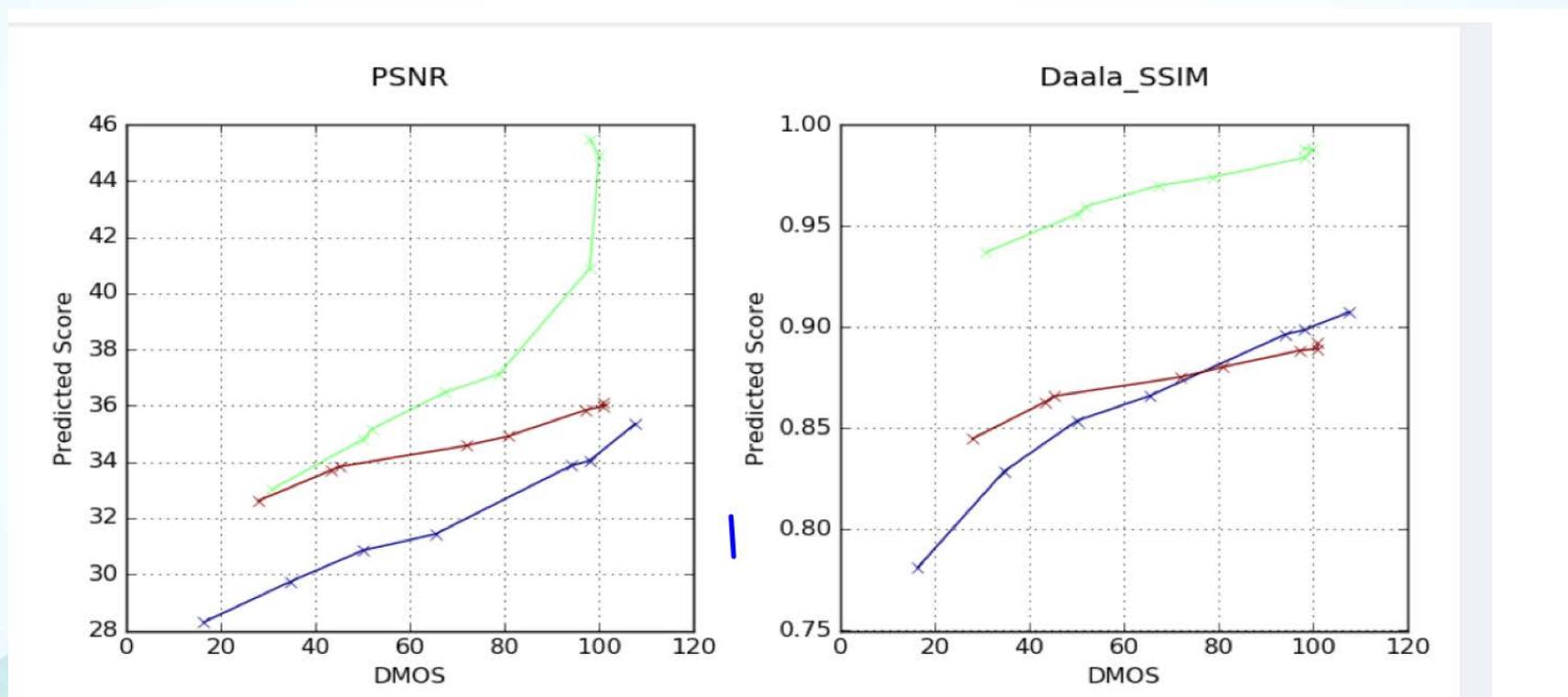
# VMAF- Netflix 的需求

上方两个视频检测出其PSNR值为大约31dB，下方两个的PSNR值约为34dB。

人们很难察觉“人群”视频有何差异，但两个“狐狸”视频的差异就很明显了。



# PSNR, SSIM, DMOS

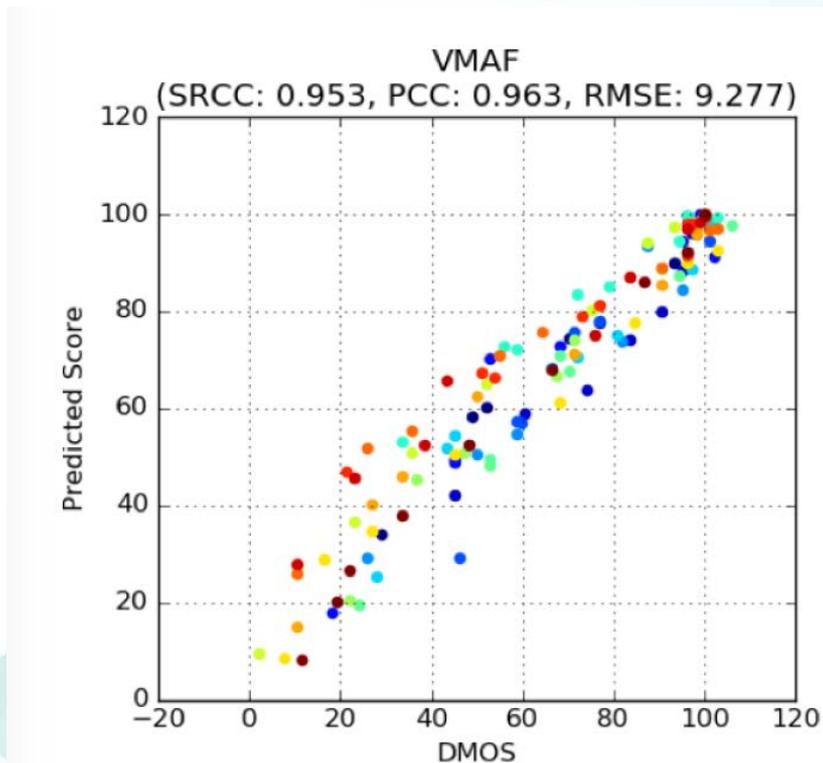
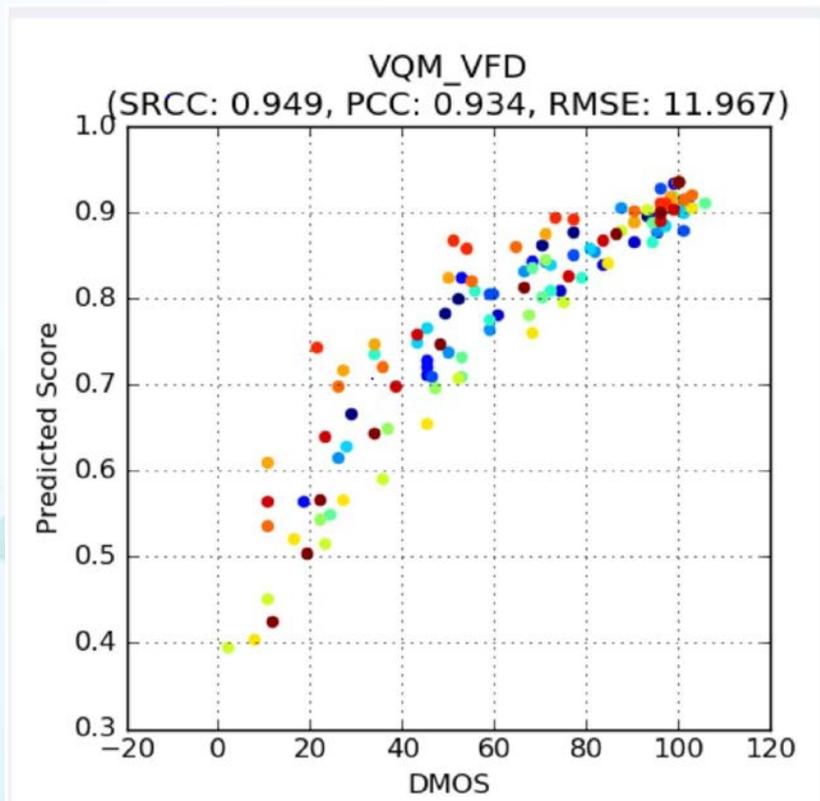


# VMAF- 技术组成

- 1) VMAF 将多种基本的质量指标结合在一起预测主观质量。
- 2) 对不同特征的源内容、失真类型，以及扭曲程度，每个基本指标各有优劣。
- 3) 通过使用机器学习算法将基本指标“融合”为一个最终指标，可以为每个基本指标分配一定的权重，这样最终得到的指标就可以保留每个基本指标的所有优势，借此可得出更精确的最终分数

# VMAF- 技术组成

- 1) 视觉信息保真度 (Visual Information Fidelity, VIF) 。
- 2) 细节丢失指标 (Detail Loss Metric, DLM)。
- 3) 运动量。这是一种衡量相邻帧之间时域差分的有效措施。



# VMAF应用：图象编码



## Haar Transforms

Originally described by A. Haar (1909). Each step creates two channels: adjacent elements (i.e., low-pass channel); and one takes difference between adjacent elements (i.e., a high-pass channel). Both are down-sampled by 2.

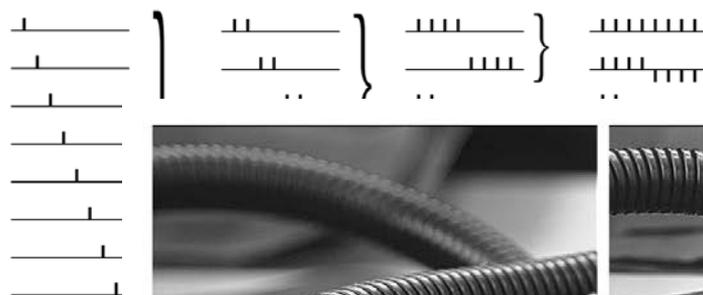
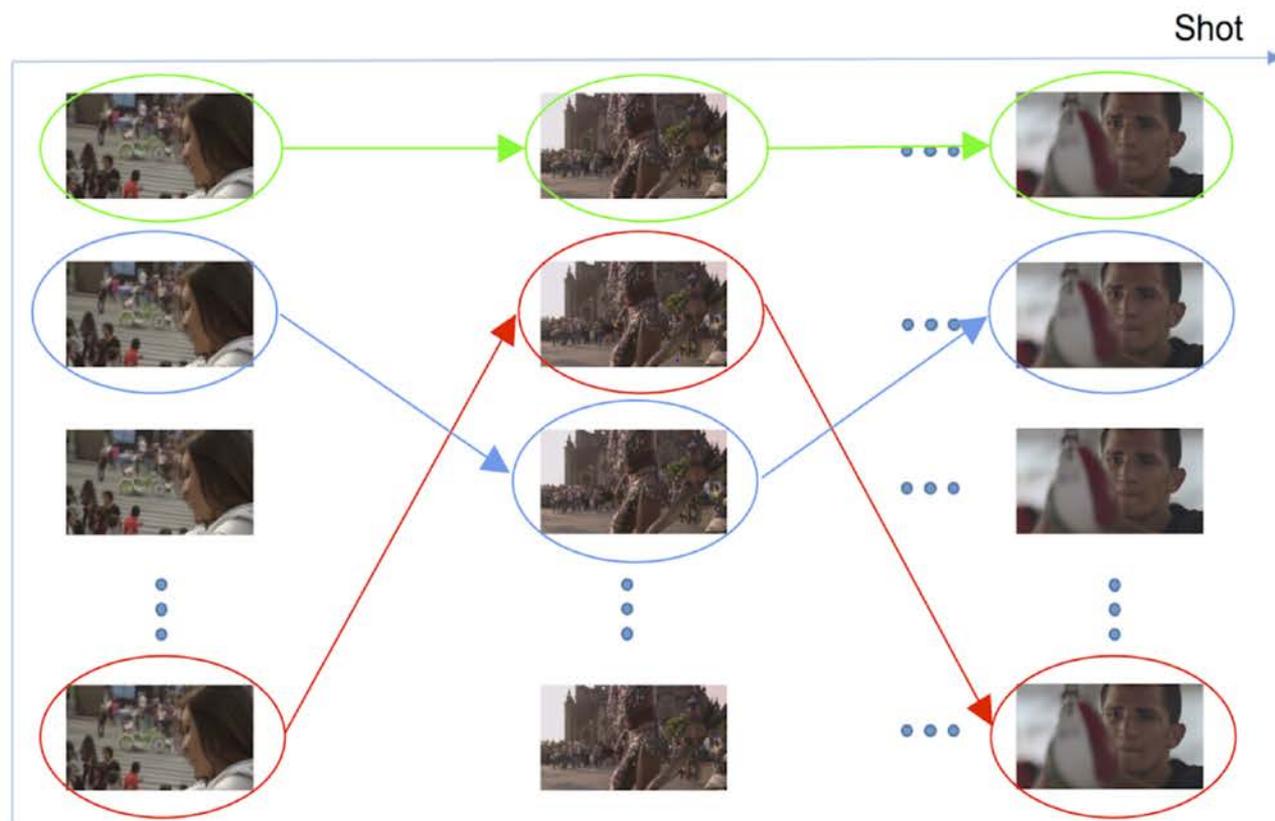


Image 1

Image 2

Composite

# VMAF应用，视频编码



Fixed QP encode

Highest (average) quality  
encode, with bitrate  $x$   
kbps

Lowest (average) bitrate  
encode, with quality  $y$

# 未来的编码解码器

- 1) VMAF 这类工具的出现, 使得多分辨率多层次的编码方案成为可能。
- 2) 多分方案收益比较大的时候, 基于物体 (pattern) 的编码就满血复活了。
- 3) aec等熵编码的应用, 使得pattern的编码效率大大提高。
- 4) pattern自适应的变换, 如AMDST, 会在更多维度上应用。比如沿着线条进行的三维AMDST, 会大大提高大尺度pattern的压缩效率。
- 5) 根据内容复杂度, 运动速度等因素自适应的动态编码会流行起来。
- 6) 机器学习会在视频编码的所有维度上展现力量。

# 参考文献

- 1) <https://medium.com/netflix-techblog/dynamic-optimizer-a-perceptual-video-encoding-optimization-framework-e19f1e3a277f>
- 2) A Bayesian Approach to Adaptive Video Super Resolution  
Ce Liu ,Microsoft Research New England.Deqing Sun,Brown University
- 3) Image super-resolution: Historical overview and future challenges  
Jianchao Yang,University of Illinois at Urbana-Champaign  
Thomas Huang,University of Illinois at Urbana-Champaign
- 4) RAISR: Rapid and Accurate Image Super Resolution, Yaniv Romano, John Isidoro, and Peyman Milanfar, Fellow, IEEE
- 5) A+: Adjusted Anchored Neighborhood Regression for Fast Super-Resolution Radu Timofte, Vincent De Smet, Luc Van Gool  
CVL, D-ITET, ETH Zürich, Switzerland, VISICS, ESAT/PSI, KU Leuven, Belgium
- 6) Image Super-Resolution via Sparse Representation,Jianchao Yang, Student Member, IEEE, John Wright, Member, IEEE Thomas Huang, Life Fellow, IEEE and Yi Ma, Senior Member, IEEE
- 7) Image Super-Resolution using Gradient Profile Prior, Jian Sun Jian Sun Zongben Xu Heung-Yeung Shum, Xi'an Jiaotong University  
Microsoft Research Asia Xi'an, P. R. China Beijing, P. R. China

# Thank You

A decorative graphic consisting of numerous colorful triangles in shades of teal, red, and green, arranged in a pattern that flows from the bottom left towards the right side of the slide.