北京航空航天大學
BEIHANG UNIVERSITY

Viewport Proposal CNN for 360° Video Quality Assessment

Mai Xu
Beihang University

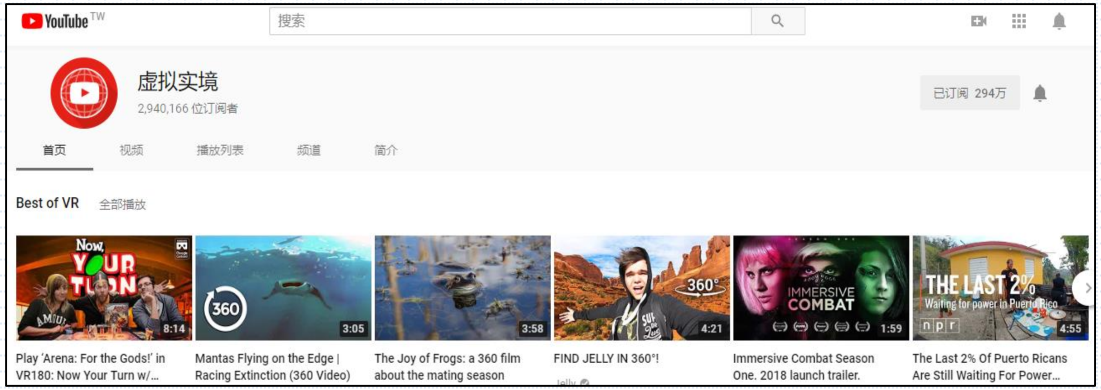
Collaborated with Chen Li, Lai Jiang, Shanyi Zhang and Xiaoming Tao

Outline

- Background
- Related Works
- Our V-CNN Approach
- Experimental Results

Background

- **Most recently, 360° video has become part of our daily life.**



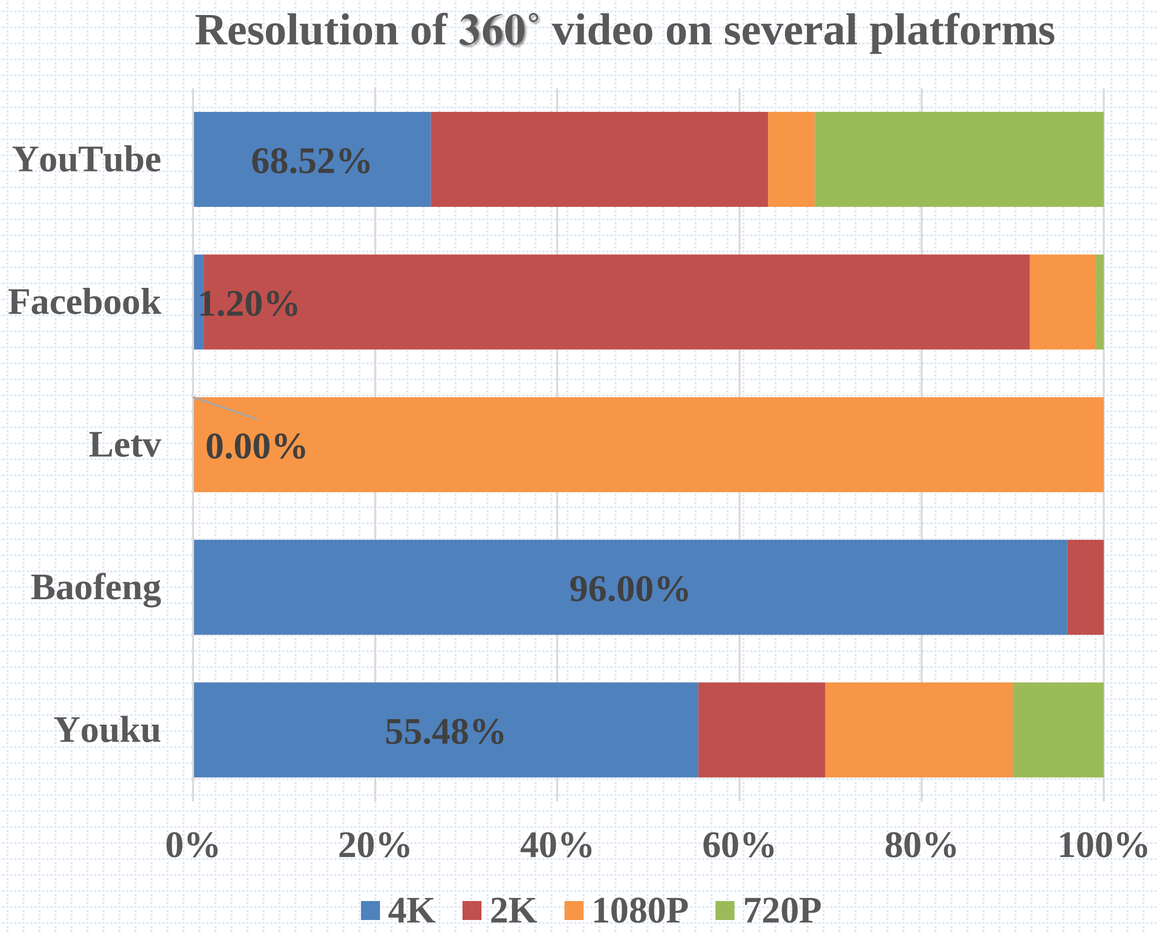
Amount of 360° video on YouTube：**100 million+**

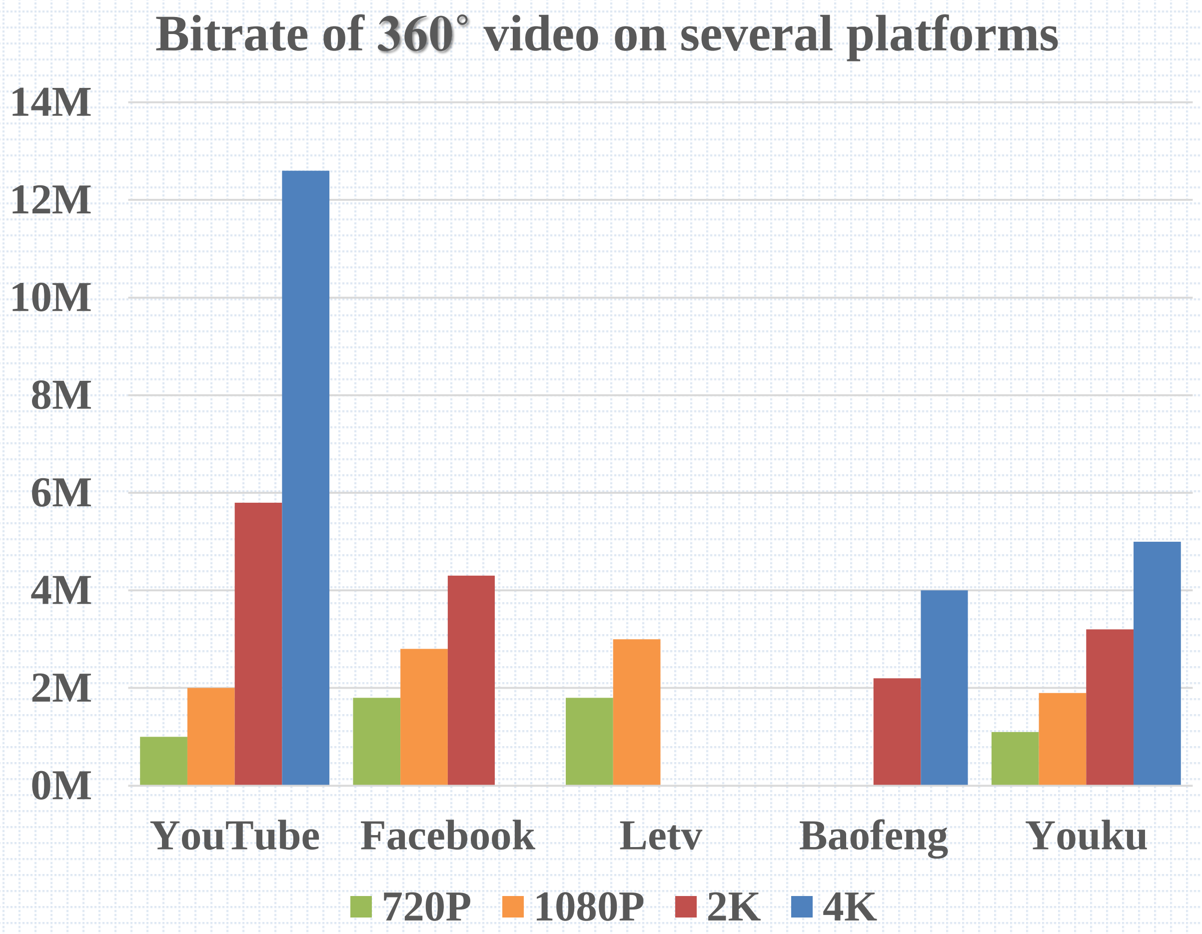
Live streaming **of CCTV Spring Festival Gala** in 360° format

Video Demo: **Skydiving**

Background

- **However, the source of 360° video varies.**





[1] Source: HUAWEI iLab, "VR data report,"2017.

Background

- **Meanwhile, the quality degradation in the viewport is more noticeable.**

Thus, it is necessary to study visual quality assessment (VQA) for 360° video.

Related Works

➢ VQA on 360° video

Non-Deep-Learning : S-PSNR (Yu *et al.*, 2015), CPP-PSNR (Zakharchenko *et al.*, 2016), WS-PSNR (Sun *et al.*, 2017), AW-SPSNR (Xiu *et al.*, 2017).

Deep-Learning: VR-IQA-NET (Lim *et al.*, 2018).

No work considers that the visual quality highly depends on the viewports in 360° video.

➢ Attention models on 360° video

HM prediction: Cheng et al., 2018, Spherical U-Net (Zhang et al., 2018), DHP (Xu et al., 2019), etc.

EM prediction: Ling et al., 2018, Xu et al., 2018, etc.

No existing 360° VQA work benefits from the auxiliary tasks of attention modelling.

➢ Great success of R-CNN approaches in object detection

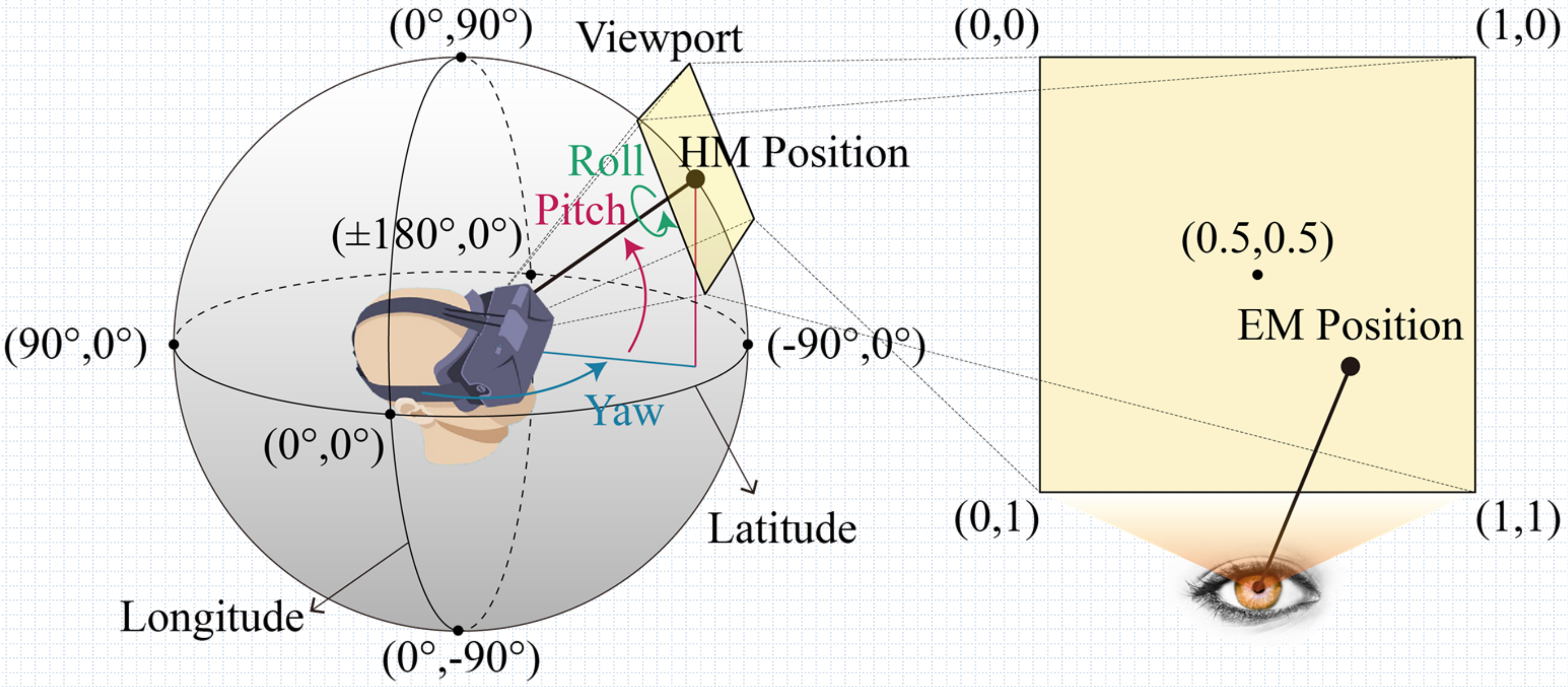
R-CNN (Girshick et al., 2014), Fast R-CNN (Girshick et al., 2015), Faster R-CNN (Ren et al., 2017), Mask R-CNN (He et al., 2017)

Two staged, multi-task approaches with region proposal.

Our V-CNN Approach

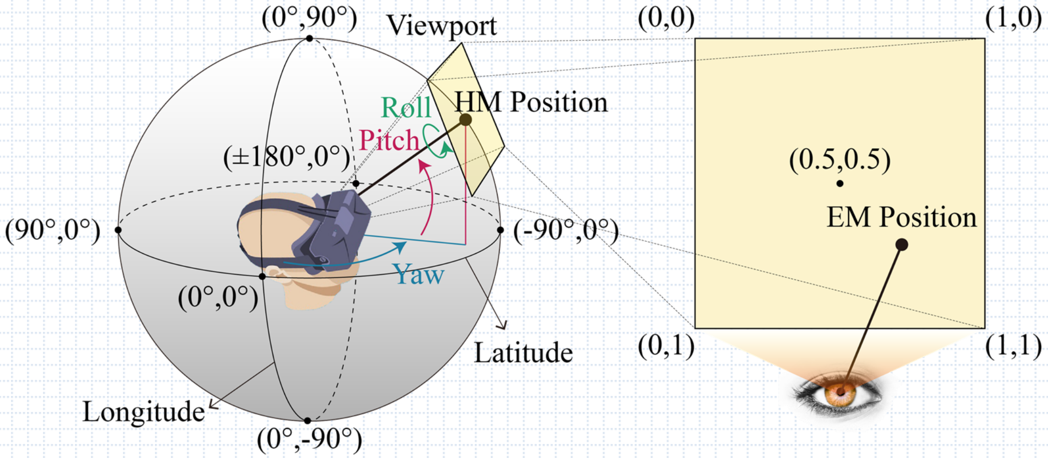
- **Motivation: Human's way of viewing 360° video:**

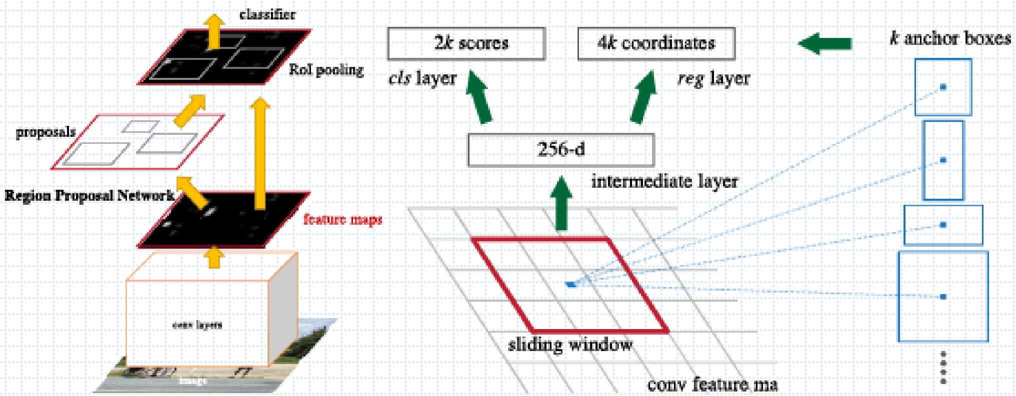
Head Movement (HM) + Eye Movement (EM)



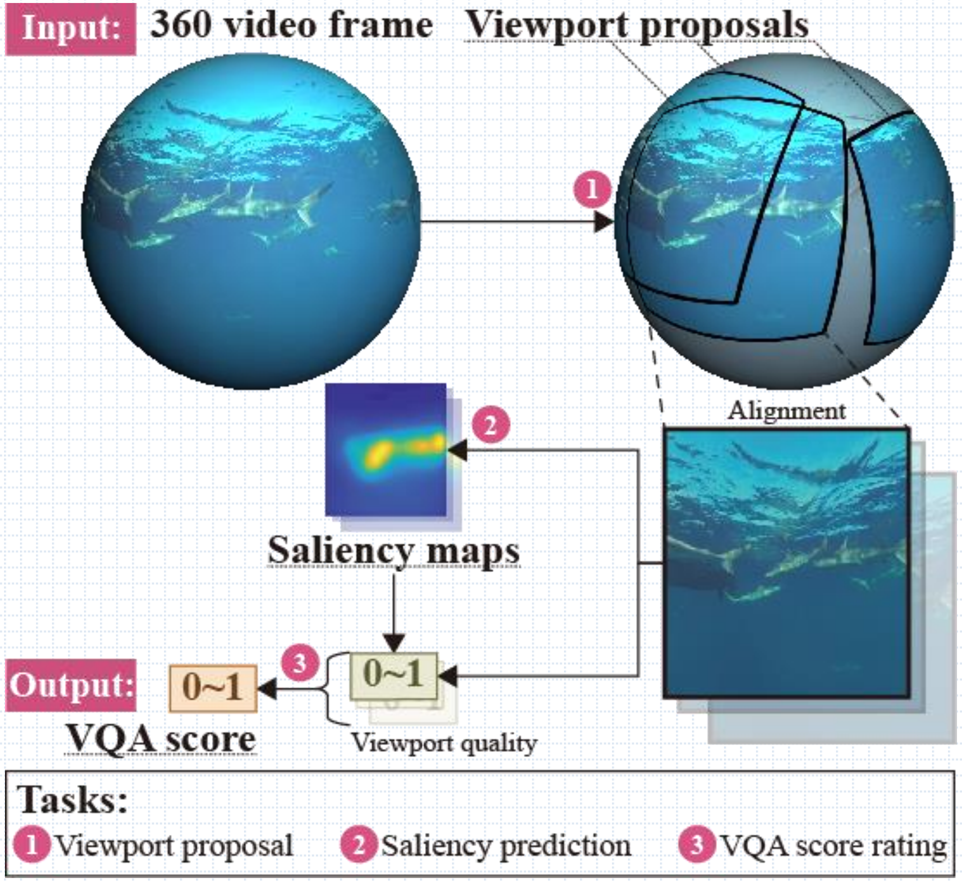
Our V-CNN Approach

- **Motivation: Two-staged procedure of viewing 360° video with viewport.**





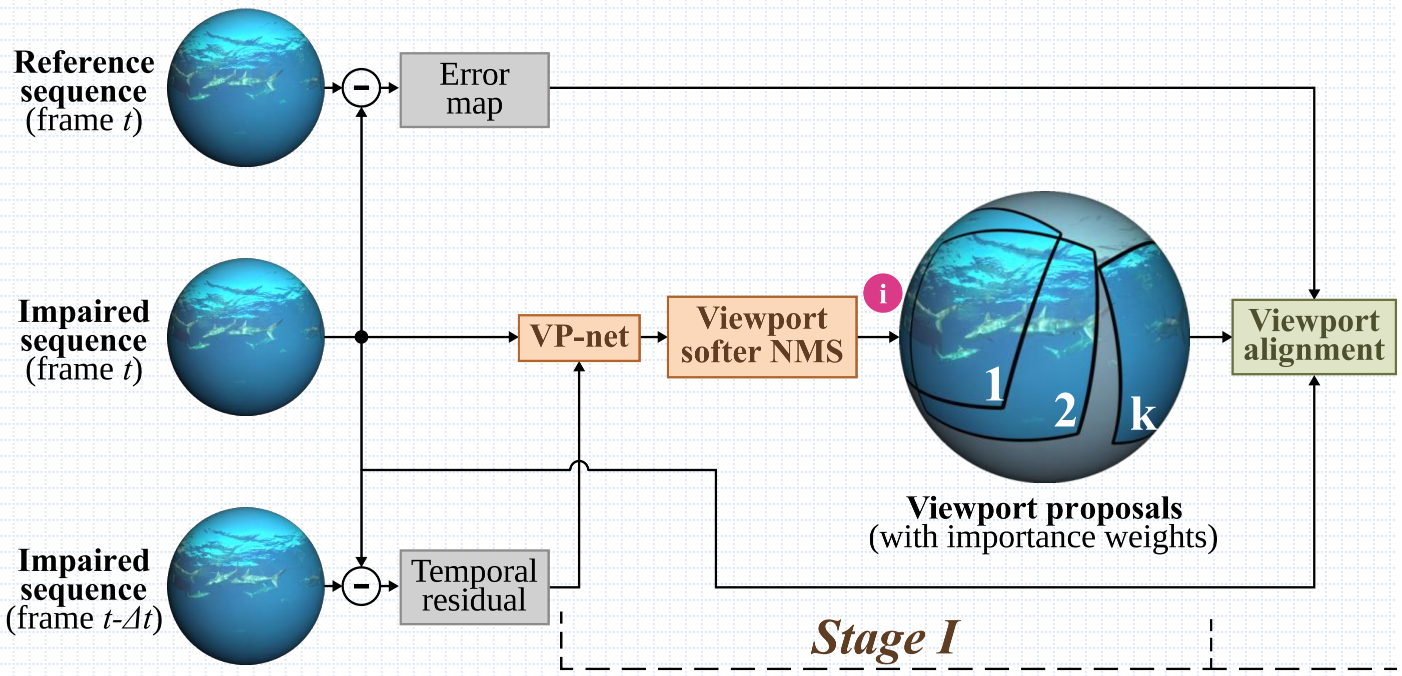
Faster R-CNN



Our V-CNN Approach

Our V-CNN Approach

- **Framework**



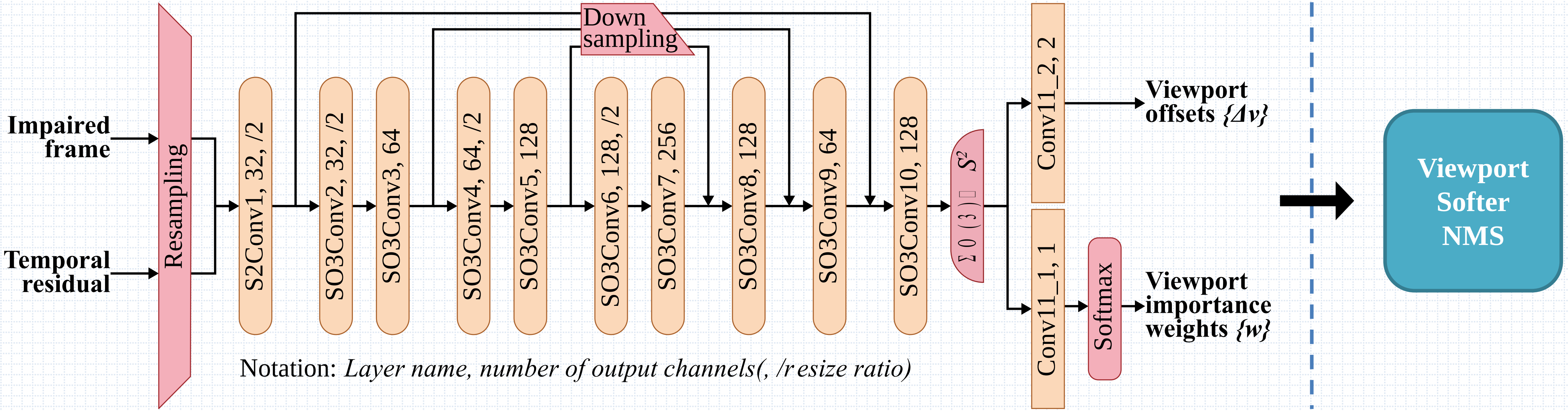
Viewport proposal

VQA

北京航空航天大学 BEIHANG UNIVERSITY

Our V-CNN Approach

- **VP-net: Viewport proposal.**



VP-Net

Loss: $\mathcal{L}^{\mathrm{I}} = \underbrace{\lambda_w \mathcal{L}_w}_{\text{Importance Weight}} + \underbrace{\lambda_{\mathbf{v}} \mathcal{L}_{\mathbf{v}}}_{\text{Anchor Location}}$

where $\mathcal{L}_w = D_{\mathrm{KL}}(\{w_i\}_{i=1}^{I} \| \{\hat{w}_i\}_{i=1}^{I}) = \sum_{i=1}^{I} w_i \log\left(\frac{w_i}{\hat{w}_i}\right)$

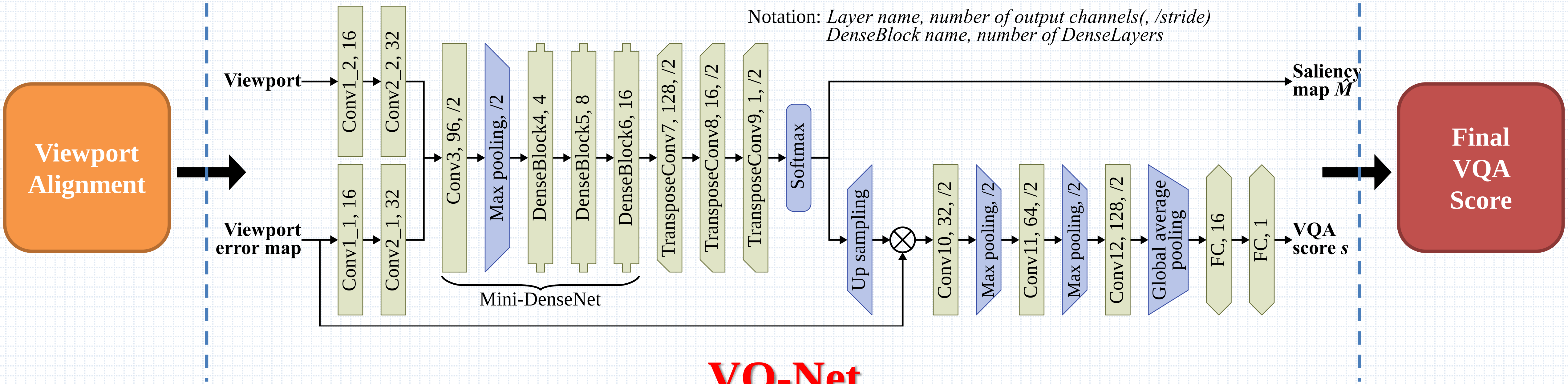
$$\mathcal{L}_{\mathbf{v}} = \sum_{i=1}^{I} w_i \cdot L_1^{\mathrm{smooth}}(\Delta\mathbf{v}_i, \Delta\mathbf{v}_i^{\mathrm{g}})$$

$$\Delta\mathbf{v}_i^{\mathrm{g}} = \arg\min_{\{\mathbf{v}_j^{\mathrm{h}}\}_{j=1}^{J}} d(\mathbf{v}_i^{\mathrm{a}}, \mathbf{v}_j^{\mathrm{h}}) - \mathbf{v}_i^{\mathrm{a}}$$

Our V-CNN Approach

北京航空航天大学 BEIHANG UNIVERSITY

- **VQ-net: Viewport quality assessment.**



VQ-Net

Loss: $\mathcal{L}_k^{\mathrm{II}} = \underbrace{\lambda_{\mathbf{M}}\mathcal{L}_{\mathbf{M}_k}}_{\text{Saliency Prediction}} + \underbrace{\lambda_s\mathcal{L}_{s_k}}_{\text{VQA Score}}$

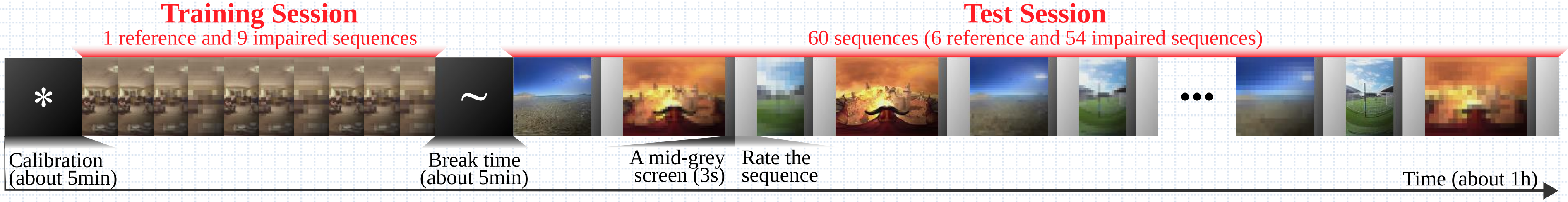
where $\mathcal{L}_{\mathbf{M}_k} = D_{\mathrm{KL}}(\mathbf{M}_k \| \hat{\mathbf{M}}_k) = \sum_{(x',y')} M_k(x',y') \log\left\{\frac{M_k(x',y')}{\hat{M}_k(x',y')}\right\}$

$\mathcal{L}_{s_k} = (s - s_k)^2$

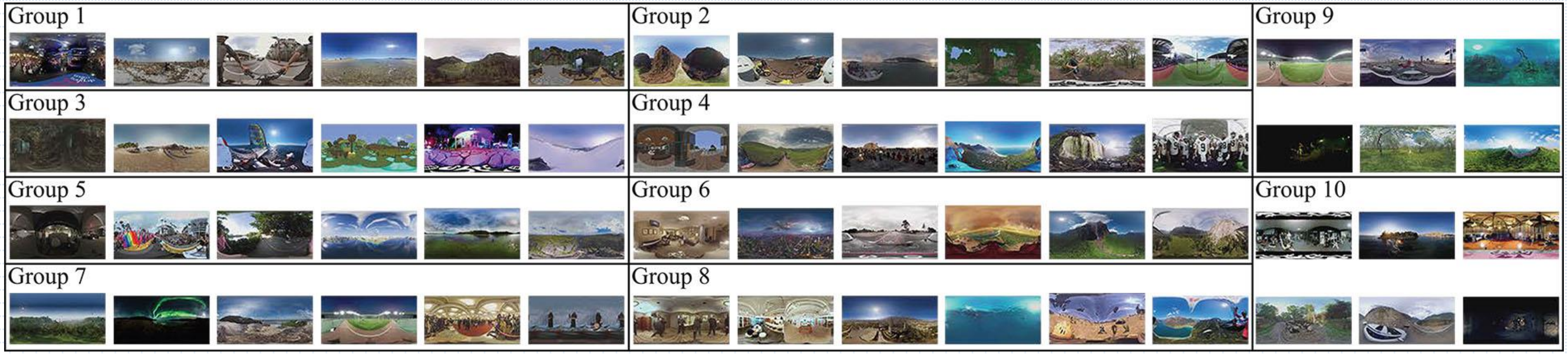
Experimental Results

▪ Dataset

☐ Collection Procedure



☐ Dataset Details



- Sequences in total: **600**
- Reference sequences: **60**
- Projection format: ERP, RCMP, TSP
- Quality level: QP = 27, 37, 42

- Subjects in total: **221**
- Including Data: Subjective quality scores, HM and EM tracking data

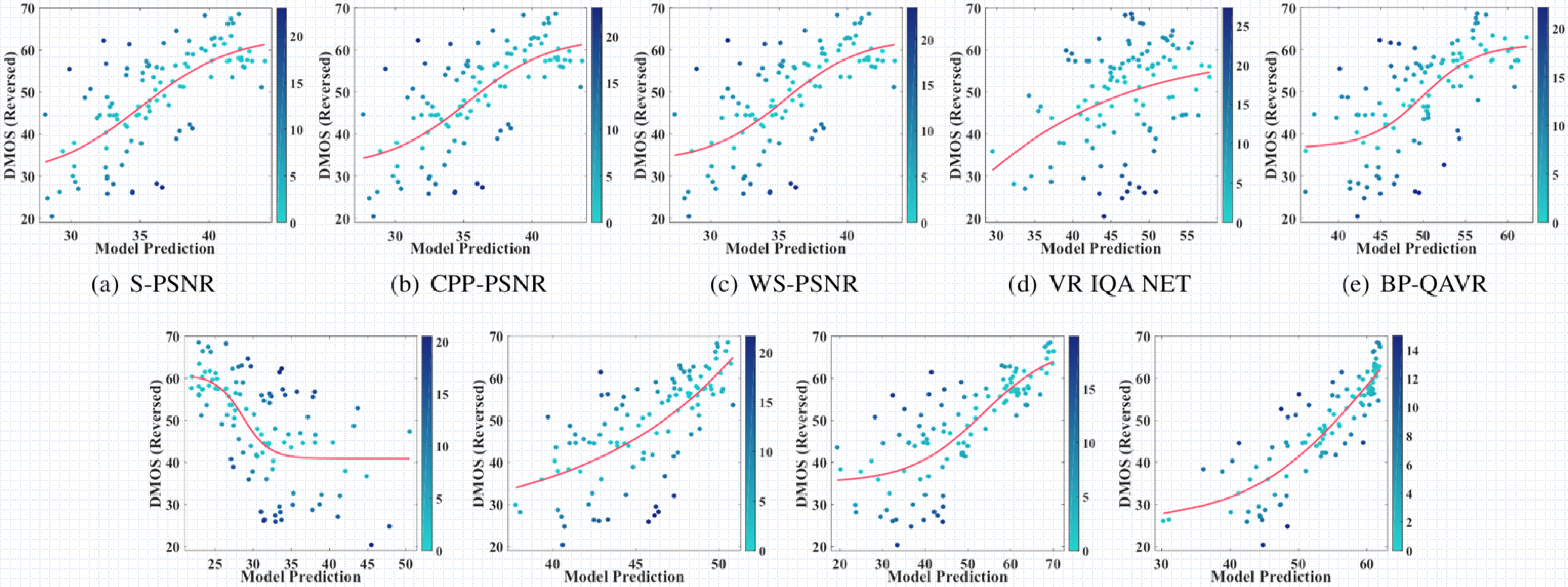
Experimental Results

- **Evaluation on VQA (Quantitative).**

Approaches	Attributes				Evaluation on VQA-ODV [19]				
	For 360° video	Full reference	Deep learning	Re-trained*	PLCC	SROCC	KROCC	RMSE	MAE
S-PSNR	✓	✓			0.6929	0.6976	0.4981	8.5407	6.6810
WS-PSNR	✓	✓			0.6721	0.6839	0.4860	8.7707	6.9089
CPP-PSNR	✓	✓			0.6812	0.6896	0.4912	8.6718	6.7932
BP-QAVR	✓	✓	✓	✓	0.6588	0.6801	0.4780	8.9112	7.0823
Li et al.	✓	✓	✓	✓	0.7821	0.7953	0.5902	7.3817	5.7793
VR-IQA-NET	✓		✓		0.3713	0.3379	0.2260	10.9984	9.1016
DeepQA		✓	✓	✓	0.6936	0.7296	0.5213	8.5325	6.7720
WaDIQaM-FR		✓	✓		0.6207	0.6162	0.4206	9.2868	7.4574
V-CNN (Ours)	✓	✓	✓	✓	**0.8740**	**0.8962**	**0.7137**	**5.7551**	**4.4893**

Experimental Results

- **Evaluation on VQA (Scatter Plots).**



(a) S-PSNR (b) CPP-PSNR (c) WS-PSNR (d) VR IQA NET (e) BP-QAVR

(f) WaDIQaM-FR (g) DeepQA (h) Li *et al.* (i) Our

Experimental Results

- **Evaluation on other auxiliary tasks.**

Viewport Proposal

Approaches	NSS	CC	KL
Cheng *et al.*	1.96	0.35	2.50
DHP	1.98	0.37	2.41
GBVS360	1.16	0.22	2.31
BMS360	1.85	0.34	**1.88**
V-CNN (Ours)	**2.65**	**0.63**	2.38

Saliency Prediction

Approaches	NSS	CC	KL
DeepVS	**1.10**	0.36	1.95
BMS	0.89	0.26	1.42
GBVS	0.81	0.30	1.34
PQFT	0.73	0.19	1.60
V-CNN (Ours)	0.97	**0.37**	**1.33**

Thank You!

Viewport Proposal CNN for 360° Video Quality Assessment

Mai Xu
Beihang University
Collaborated with Chen Li, Lai Jiang, Shanyi Zhang and Xiaoming Tao