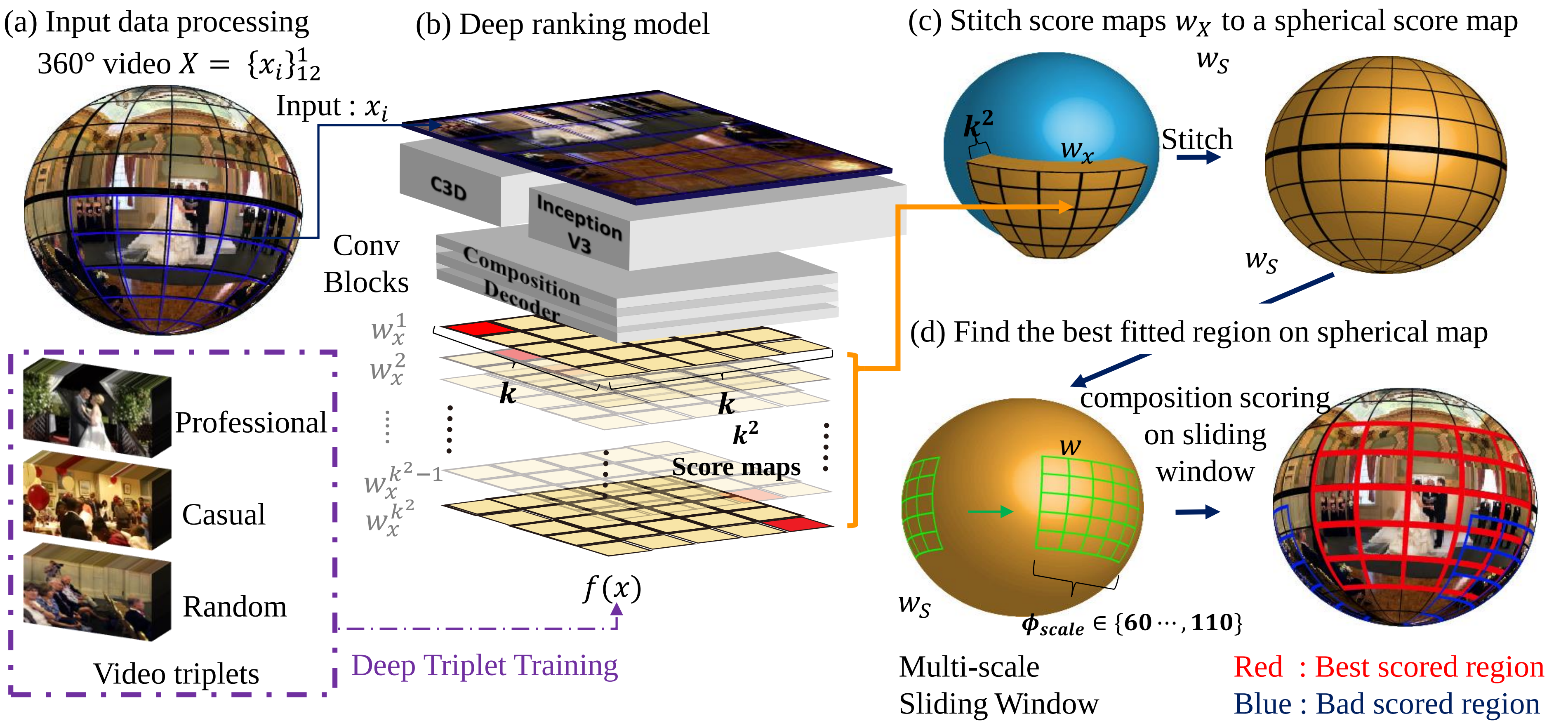
A Deep Ranking Model for Spatio-Temporal Highlight Detection from a 360º Video



(a) Input data processing

360° video $X = \{x_i\}_{12}^{1}$

(b) Deep ranking model

(c) Stitch score maps w_X to a spherical score map w_S

(d) Find the best fitted region on spherical map

Multi-scale
Sliding Window

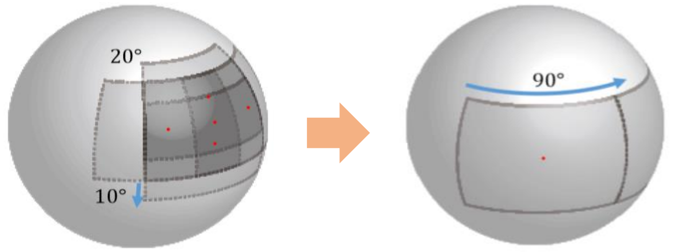
Red : Best scored region
Blue : Bad scored region

(1) **Fully convolutional CVS** generates a layered spherical score maps.

(2) **Position-wise composition score** learns fidelity of views and determines which view is suitable for highlight.

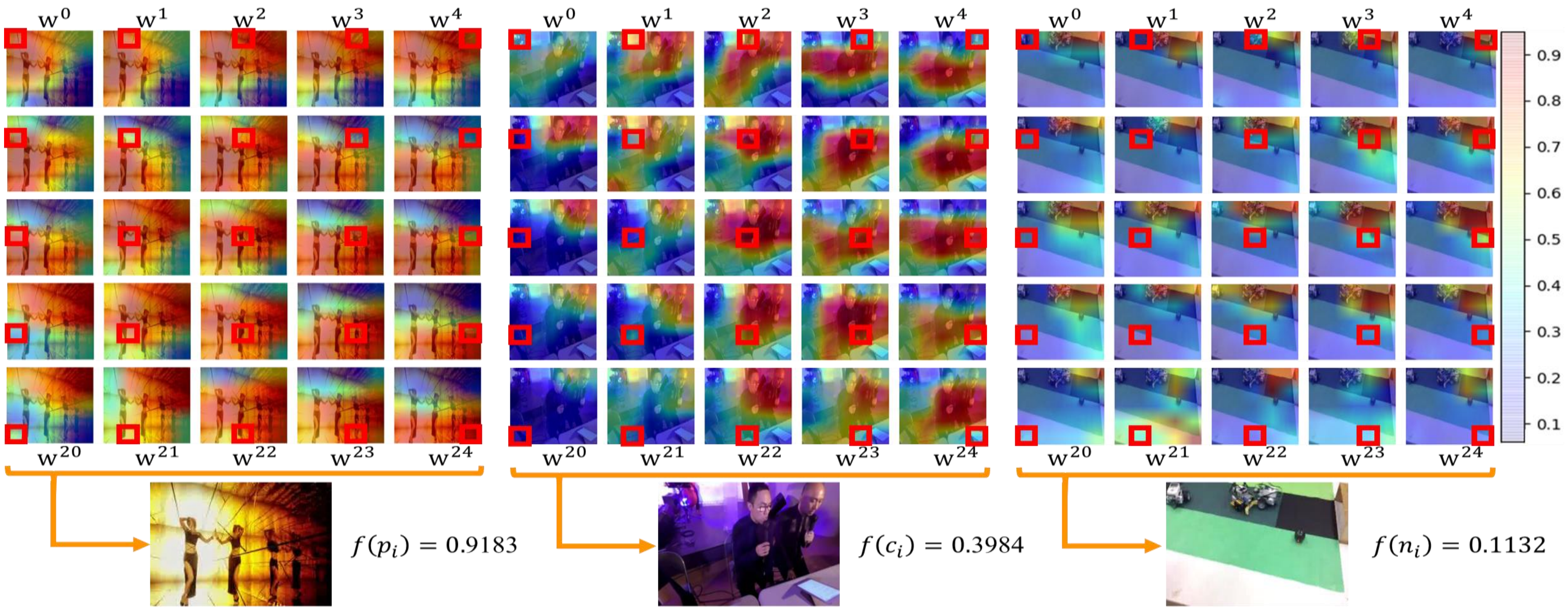
(3) **Reduces a bottleneck** of normal field-of-view projection

AutoCam (Su et al. 2016) Ours (CVS)



# of glimpse 198 patches >> **12 patches**

Composition Score Map Learned by Triplet Deep Ranking

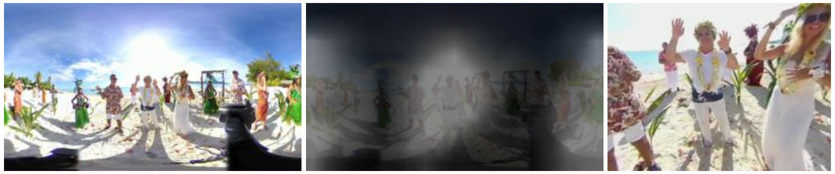


(a) Professional

(b) Casual

(c) Random

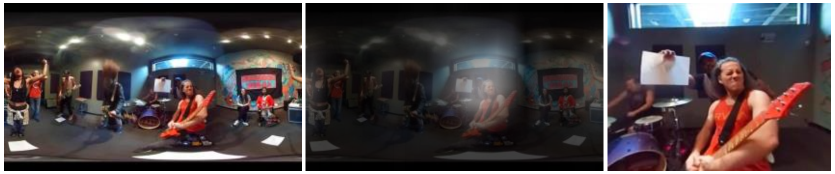
$$f(p_i) \succ f(c_i) \succ f(n_i), \quad \forall (p_i, c_i, n_i) \in \mathcal{D}$$



Original

Score map

Projection



Original

Score map

Projection

Youngjae Yu

Sangho Lee

Joonil Na

Jaeyun Kang

Gunhee Kim

Association for the Advancement of Artificial Intelligence

SEOUL NATIONAL UNIV. VISION & LEARNING